DATA-BALANCED CURRICULUM LEARNING FOR AUDIO QUESTION ANSWERING

Technical Report

Gijs Wijngaard[1], *Elia Formisano*[2,3,4], *Michele Esposito*[1], *and Michel Dumontier*[1]

[1]Department of Advanced Computing Sciences, Maastricht University
[2]Department of Cognitive Neuroscience, Maastricht University
[3]Maastricht Centre for Systems Biology, Maastricht University
[4]Brightlands Institute for Smart Society, Maastricht University

ABSTRACT

Audio question answering (AQA) requires models to understand acoustic content and perform complex reasoning. Current models struggle with dataset imbalances and unstable training dynamics. This work combines curriculum learning with statistical data balancing to address these challenges. The method labels question difficulty using language models, then trains progressively from easy to hard examples. Statistical filtering removes overrepresented audio categories, and guided decoding constrains outputs to valid multiple-choice formats. Experiments on the DCASE 2025 training set and five additional public datasets show that data curation improves accuracy by 19.2% over baseline models, achieving 64.2% on the DCASE 2025 benchmark.

1. INTRODUCTION

Audio Question Answering (AQA) represents a fundamental challenge at the intersection of acoustic understanding and natural language processing. Unlike simple sound classification, AQA requires models to comprehend complex acoustic scenes, identify temporal relationships between sounds, and generate coherent answers to diverse questions [1, 2, 3]. Where earlier AQA models focused on binary tasks or limited label sets [1, 4], later models contain answer generation capabilities in natural language beyond strict answer sets. Some tasks require text generation metrics to evaluate the output of the model, while other tasks require selecting the correct answer from provided options, which makes answers verifiable against actual answers.

Audio-language modeling faces limited diversity and quality of training data [5]. Many models train on overlapping datasets, and this redundancy also appears in benchmarks in the form of data contamination, which contributes to homogeneous dataset landscapes. This limits the generalization and robustness of audio-language models. Analysis reveals that audio datasets suffer from severe class imbalances. Certain sound categories show dramatic over-representation. This leads to models that perform well on common sounds but fail on rare acoustic events.

While traditional approaches to audio question answering have relied on supervised learning, recent advances have explored reinforcement learning techniques [6]. Reinforcement learning offers potential advantages for complex reasoning tasks but introduces training instabilities. This work investigates how curriculum learning and guided decoding can stabilize reinforcement learning for audio understanding.

This work tests the following hypotheses for improving audio question answering:

1. **Curriculum-guided reinforcement learning**: The model trains on easy examples first to establish reliable reward signals. Difficult samples enter training progressively. This method stabilizes learning dynamics.
2. **Statistical data balancing**: Statistical thresholds identify and remove overrepresented categories, which balances the training datasets.
3. **Guided decoding**: Regular expressions constrain generation to valid multiple-choice answers (A, B, C, D).
4. **Hybrid training**: The model trains with Supervised Fine-Tuning (SFT) to provide stable initialization. Afterwards, the training paradigm uses Group Relative Policy Optimization (GRPO) through reward-based learning.

Experiments on six datasets validate the approach. Data quality determines performance more than algorithmic complexity. The method achieves 64.2% accuracy on DCASE 2025 Task 5: Audio Question Answering.

2. BACKGROUND

2.1. Audio-Language Models

Large-scale audio-language models transform audio understanding. Early approaches combine audio encoder models as encoder with language models as decoder [7, 8]. Here, the language model's performance is dependent on the quality of the features it receives from the audio encoder. Qwen2-Audio [9] and other recent models address this through unified architectures. The model processes a combination of speech, audio, and environmental sounds with a Whisper-large-v3 encoder and a Qwen-7B language model, and is used in work as a foundational model.

Examples of training strategies in the audio-language domain include: temporal progression from 30-second to 5-minute contexts that enable 3B models to outperform larger architectures [10]. Perception-before-understanding improves comprehension [11]. Multi-phase thinking, to allow planning, captioning, reasoning, and summarization [12].

This work was supported by the Dutch Research Council (NWO 406.20.GO.030 to Prof. Elia Formisano), the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-12157, Data Science Research Infrastructure (DSRI; Maastricht University) and the Dutch Province of Limburg.

2.2. Audio Question Answering Datasets

Challenge rules from the DCASE Challenge restrict dataset selection to 20 approved sources. Many datasets target other tasks such as audio captioning and audio generation rather than question answering. This work selects only datasets with multiple-choice formats or relevant question-answer pairs. Five extra datasets are used to train the models, next to the training set of the challenge:

- **AVQA** [13]: Contains 56,369 question-answer pairs from VGGSound (originally 57,335 before YouTube deletions). Questions span eight semantic types. The dataset includes audio and visual information. AVQA requires models to discover causal correlations between modalities.
- **ClothoAQA** [14]: Contains 35,838 questions from 1,991 environmental audio samples (15-30 seconds). Each audio has 6 questions with 3 annotator responses per question. Four questions require binary answers; two require single words. Different annotators create questions and answers.
- **CompA-Order** [15]: Contains 900 examples for compositional reasoning. The dataset tests temporal understanding between sounds.
- **TACOS** [16]: Contains 12k audio recordings with 61,137 temporally-aligned captions. The dataset trains temporal reasoning.
- **AudSem** [17]: Synthesized from YouTube closed captions. Contains 210,288 multiple-choice questions across diverse acoustic scenes and events.

Evaluation Datasets. The DCASE 2025 Challenge provides three specialized evaluation datasets. **Part 1 (Bioacoustics QA)** contains 0.7K training and 0.2K development examples from 31 marine mammal species with sample rates from 600 Hz to 160 kHz from the Watkins Marine Mammal Sound Database [18]. Audio duration spans 0.4 to 625 seconds. **Part 2 (Temporal Soundscapes QA)** includes 1K training and 0.6K development examples covering 26 sound classes. The 10-second mono audio files (32-48 kHz) test temporal reasoning about sound sequences, timestamps, and durations. Part 2 sources include NIGENS general sound events database [19], L3DAS23 Challenge [20], and TAU Spatial Sound Events 2019 [21]. **Part 3 (Complex QA)** contains 6.4K training and 1.6K development examples from AudioSet [22] and Mira datasets [23]. Each 10-second audio clip (16 kHz) requires multi-faceted reasoning across temporal, acoustic, and contextual dimensions.

3. METHOD

The method combines four components: (1) Curriculum-based filtering controls learning progression. (2) Statistical balancing ensures dataset diversity. (3) Hybrid training combines supervised fine-tuning with reinforcement learning. (4) Guided decoding for structuring output generation during GRPO training.

3.1. Curriculum-Based Data Filtering

Traditional curation removes difficult examples as noise. These samples provide valuable signals when introduced properly. Curriculum learning trains models on easy examples before difficult ones [24, 25]. To be able to get a subset of the dataset of easy examples to train the model on first, an LLM is used to label the dataset. Here, Microsoft's Phi-4-mini-instruct [26] scores question difficulty from 0.0 (very easy) to 1.0 (very difficult). For each example (a_i, q_i, c_i, y_i) in the dataset, the model evaluates: question complexity, required knowledge depth, ambiguity level, and concept count.

After labeling, the samples are filtered by difficulty. Easy samples (difficulty < 0.3) establish reliable GRPO reward signals, which prevents instability from ambiguous examples. Later stages incorporate all samples or focus on hard examples (difficulty > 0.7). Best results come from training on easy samples, then the full dataset.

3.2. Statistical Category Balancing

Audio datasets exhibit severe category imbalances. Dataset curation addresses class imbalances in audio datasets. AudioSet contains 527 classes with severe imbalances: Music and Speech dominate [22]. A language model assesses caption quality against a description of human standards, where low-scoring captions get filtered [27]. Analysis identifies 24 audio categories: speech (conversation and monologue), music, nature sounds, urban sounds, transportation, and specialized domains. Music and speech dominate with thousands of examples. Specialized sounds have dozens.

Phi-4-mini-instruct [26] categorizes audio questions into 24 types. The model infers categories from question content using 256-sample batches. In total, we have 365,479 sounds across 24 audio categories.

In the filtering process, statistical thresholding balances category distributions. The method computes mean count μ across categories, and categories with count $c_i > \theta \cdot \mu$ get filtered. Optimal performance occurs at $\theta = 0.7$. This threshold removes extreme over-representation of human sounds and mixed environment sounds while maintaining sufficient examples. The approach prevents overfitting to dominant categories. In Figure 1, the audio categories are visualized with their respective count and the cut-off threshold. The figure shows that human sounds and mixed environment sounds are over-represented in the training sets.

3.3. Hybrid SFT-GRPO Training Pipeline

The training pipeline combines supervised fine-tuning and reinforcement learning. Group Relative Policy Optimization (GRPO) [28] is a training technique introduced recently for reward-based optimization in finetuning models such as DeepSeek R1 [28] and Qwen2.5 [29]. Supervised Fine-Tuning (SFT) works well as a warm-up stage before GRPO, to accustom the model to the proposed formatting and style. Qwen2-Audio serves as the base model.

Stage 1: Supervised Fine-Tuning (SFT). The method finetunes Qwen2-Audio-7B-Instruct using cross-entropy loss. Low-Rank Adaptation (LoRA) [30] targets query and value projections with rank $r = 8$, scaling factor $\alpha = 16$, and dropout 0.05. Training uses AdamW optimizer [31], learning rate 2×10^{-5}, cosine annealing and gradient clipping at 0.5. Training runs 1-3 epochs.

Stage 2: Group Relative Policy Optimization (GRPO). GRPO optimizes performance through reinforcement learning with verifiable rewards (RLVR) after SFT convergence. The reward function combines accuracy and format validation:

$$R(y, \hat{y}) = R_{acc}(y, \hat{y}) + R_{format}(\hat{y})$$

R_{acc} awards 0.5 for full match, 0.25 for matching the answer letter (e.g. A, B, C, or D), and 0.25 for remaining content. R_{format} awards 0.5 for correct format. GRPO uses $\beta = 0.01$ with a warmup

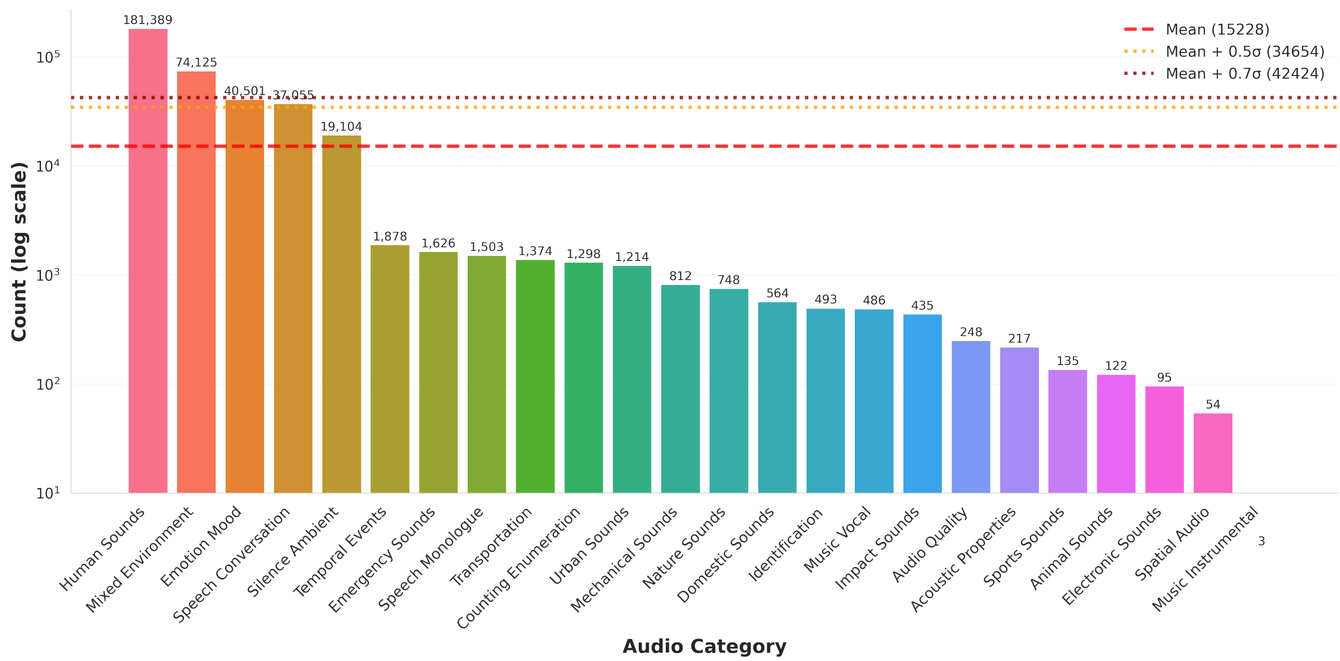

Figure 1: Distribution of audio categories in the training sets. Dotted lines represent the threshold during diversity filtering, with the cut-offs being 0.5σ and 0.7σ above the mean.

of 50 steps, 4 generations per example, for 1-2 epochs, depending on the configuration. Dropout is disabled and the learning rate is 1×10^{-6} with AdamW with cosine annealing.

Stage 3: Multi-Stage Pipeline. Effective training chains three stages: (1) SFT on all datasets with LoRA. (2) GRPO on easy samples with diversity balancing. (3) GRPO on the full dataset with diversity balancing. Each stage builds on previous checkpoints. The pipeline combines SFT initialization with progressive GRPO improvements.

3.4. Guided Decoding with Constrained Outputs

Models must produce valid multiple-choice answers. Guided decoding constrains a language model's output to ensure it conforms to a specific structure, using a regular expression [32]. This is achieved by compiling the desired structure into a finite state machine, which can be visualized as a graph where each node represents a valid state in the grammar. At each step of text generation, the model can only transition to a new state (i.e., generate a next token) if there is a valid edge from its current position in the graph. This process works by creating a "logit mask" that effectively sets the probability of all invalid next tokens to zero, forcing the model to choose only from the tokens that keep the output syntactically correct according to the graph's rules. Constrained decoding ensures correct output format during generation.

Regular expressions constrain each decoding step. Valid paths correspond to answers A, B, C, or D. The regular expression `^<think>.*?</think>\s*<answer>(A|B|C|D).*</answer>$` restricts the model on generation for choosing an answer, while also allowing for reasoning. This approach of guided decoding eliminates post-processing since mathematical constraints force selection from four answer choices.

Evaluation uses the DCASE 2025 evaluation audio question-answering datasets. Experiments compare against baseline models and conduct ablation studies. In total, 22 experimental runs are conducted comparing different training strategies.

3.5. Experimental Setup

Datasets: AVQA (56,369 QA pairs), ClothoAQA (35,838 questions), CompA-Order (900 examples), TACOS (61,137 captions), and AudSem (210,288 questions). **Baselines**: Qwen2-Audio, AudioFlamingo 2, and Gemini-2.0-Flash. **Metrics**: Top-1 accuracy with exact match evaluation. Results report overall accuracy and part-wise performance (Part 1: bioacoustics, Part 2: temporal/counting, Part 3: complex reasoning).

4. RESULTS

Table 1 presents the evaluation results on the DCASE 2025 test set submitted to the DCASE Challenge. The four training configurations are:

- **SFT only**: A baseline model trained with supervised fine-tuning on all datasets for one epoch using LoRA optimization with rank 8 and learning rate 2×10^{-5}.
- **GRPO + Curriculum**: A three-stage pipeline that first performs SFT, then GRPO training on easy samples (difficulty < 0.3) with diversity balancing at threshold 0.5, followed by GRPO on the full DCASE2025 training dataset for 5 epochs with guided decoding constraints.
- **GRPO + Div.** ($\theta = 0.7$): Combines 3-epoch SFT training with aggressive diversity filtering that removes overrepresented categories when their count exceeds 0.7 times the mean category

Model	Part 1	Part 2	Part 3	Total
Baseline Models				
Qwen2-Audio-7B	30.0%	39.2%	49.6%	45.0%
AudioFlamingo2	53.9%	31.7%	49.5%	45.7%
Gemini-2.0-Flash	42.0%	46.3%	56.6%	52.5%
Our Methods				
SFT only	62.1%	**42.2%**	72.5%	64.1%
GRPO + Curriculum	67.0%	38.3%	**72.8%**	63.7%
GRPO + Div. ($\theta = 0.7$)	65.6%	41.4%	72.6%	**64.2%**
Ensemble (11 models)	**75.0%**	**42.2%**	68.3%	62.5%

Table 1: Accuracy on the DCASE 2025 evaluation set. Best results per column in bold.

count, followed by two GRPO stages with the same diversity threshold.

- **Ensemble**: A majority voting ensemble combining all three individual models plus eight additional variants trained with different curriculum strategies and diversity thresholds.

Part 2 (temporal and counting tasks) remains the most challenging with accuracies between 38-42%, which indicates fundamental difficulties in temporal reasoning within audio-language models. Part 1 (bioacoustics) and Part 3 (complex reasoning) show stronger performance, achieving 75.0% and 72.8% in the best scoring models respectively.

4.1. Ablation of Model Components

Configuration	Accuracy
Training Paradigm	
Pre-trained Qwen2-Audio (no fine-tuning)	45.0%
SFT only (single dataset)	64.0%
SFT only (all datasets)	63.9%
SFT + GRPO	63.1%
Diversity Balancing	
No diversity filtering	63.1%
Diversity $\theta = 0.3$	63.2%
Diversity $\theta = 0.7$	64.2%
Curriculum Learning	
No curriculum	63.1%
Easy samples (difficulty < 0.2)	64.2%
Hard samples (difficulty > 0.7)	63.1%
Additional Techniques	
LoRA in GRPO	62.3%
Guided decoding	63.7%

Table 2: Ablation study results showing the impact of different components on DCASE 2025 test set accuracy. All ablations are built upon the SFT+GRPO training pipeline, which serves as our internal baseline (63.1%).

Table 2 summarizes an ablation study isolating the impact of different training strategies and data curation techniques.

Supervised fine-tuning with LoRA optimization forms the foundation of this approach. The pre-trained Qwen2-Audio model achieves 45.0% accuracy on DCASE 2025. SFT with LoRA on a single dataset increases accuracy to 64.0%. Expanding to all six datasets yields 63.9% accuracy, which indicates dataset diversity alone provides minimal improvement without proper curation.

The multi-stage SFT+GRPO approach maintains 63.1% accuracy, showing modest benefits compared to SFT alone. Notably, the ablation results reveal that GRPO does not outperform SFT when used alone, suggesting that reinforcement learning's effectiveness is conditional on data curation techniques that provide cleaner, more stable reward signals. Curriculum learning on easier examples (difficulty < 0.2) achieves 64.2%, while focusing on hard examples (difficulty > 0.7) reduces performance to 63.1%.

Diversity balancing produces the most significant improvements. Conservative filtering ($\theta = 0.3$) yields 63.2% accuracy, while aggressive filtering ($\theta = 0.7$) achieves the best single-model result of 64.2%. This confirms that removing overrepresented audio categories contributes more than simply adding data. Guided decoding experiments show modest improvement, as the model naturally learns to generate well-formatted outputs during training.

4.2. Reasoning Phase

Incorporating a reasoning phase is also tried, where models generate intermediate reasoning steps before producing final answers. It yields significantly lower results (46.2-51.6% accuracy) compared to the best model's 64.2%. This limitation stems from dataset constraints: only AudSem among the six datasets includes thinking annotations. While GRPO theoretically enables thinking across all datasets by evaluating only final answers, poor SFT initialization from limited thinking examples prevents effective learning. Thinking-based approaches require comprehensive annotated reasoning across diverse datasets to compete with direct answer generation methods.

5. CONCLUSION

The method makes three contributions: (1) Statistical diversity balancing with threshold 0.7 addresses dataset imbalances. (2) Curriculum learning improves specific question types without significant overall gains. (3) Supervised fine-tuning with LoRA optimization delivers better performance than reinforcement learning alone, with SFT outperforming the SFT+GRPO approach unless combined with proper data curation techniques.

GRPO limitations highlight fundamental differences between audio question answering and text generation tasks. Future work requires alternative reward formulations or denser feedback mechanisms. The success of diversity filtering motivates more sophisticated curation approaches using learned representations instead of predefined categories.

Dataset quality matters more than algorithmic complexity in audio question answering. Data curation outperforms complex algorithms for audio question answering. Diversity filtering provides the largest performance gains, achieving 64.2% accuracy on DCASE 2025 Task 5. Diversity filtering applies to any audio-language dataset with minimal computational overhead. Balanced acoustic representation ensures more fair real-world performance across sound categories.

6. REFERENCES

[1] G. Li, Y. Xu, and D. Hu, "Multi-scale attention for audio question answering," in *Proc. Interspeech 2023*, 2023, pp. 3442–3446.

[2] S. R. Behera, P. B. Reddy, A. M. Tripathi, M. B. Rathod, and T. Karavadi, "Towards Multi-Lingual Audio Question Answering," in *INTERSPEECH 2023*. ISCA, Aug. 2023, pp. 356–360.

[3] H. M. Fayek and J. Johnson, "Temporal Reasoning via Audio Question Answering," *IEEE ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2283–2294, 2020.

[4] P. Sudarsanam and T. Virtanen, "Attention-Based Methods For Audio Question Answering," in *31st European Signal Processing Conference, EUSIPCO 2023, Helsinki, Finland, September 4-8, 2023*. IEEE, 2023, pp. 750–754.

[5] G. Wijngaard, E. Formisano, M. Esposito, and M. Dumontier, "Audio-Language Datasets of Scenes and Events: A Survey," *IEEE Access: Practical Innovations, Open Solutions*, vol. 13, pp. 20 328–20 360, 2025.

[6] G. Li, J. Liu, H. Dinkel, Y. Niu, J. Zhang, and J. Luan, "Reinforcement Learning Outperforms Supervised Fine-Tuning: A Case Study on Audio Question Answering," Mar. 2025.

[7] X. Mei, X. Liu, Q. Huang, M. D. Plumbley, and W. Wang, "Audio Captioning Transformer," Jul. 2021.

[8] Y. Koizumi, R. Masumura, K. Nishida, M. Yasuda, and S. Saito, "A transformer-based audio captioning model with keyword estimation," in *Proc. Interspeech 2020*, 2020, pp. 1977–1981.

[9] Y. Chu, J. Xu, Q. Yang, H. Wei, X. Wei, Z. Guo, Y. Leng *et al.*, "Qwen2-Audio Technical Report," Jul. 2024.

[10] S. Ghosh, Z. Kong, S. Kumar, S. Sakshi, J. Kim, W. Ping, R. Valle *et al.*, "Audio Flamingo 2: An Audio-Language Model with Long-Audio Understanding and Expert Reasoning Abilities," Mar. 2025.

[11] Y. Gong, H. Luo, A. H. Liu, L. Karlinsky, and J. R. Glass, "Listen, think, and understand," in *The Twelfth International Conference on Learning Representations*, 2024.

[12] Z. Xie, M. Lin, Z. Liu, P. Wu, S. Yan, and C. Miao, "Audio-Reasoner: Improving Reasoning Capability in Large Audio Language Models," Mar. 2025.

[13] P. Yang, X. Wang, X. Duan, H. Chen, R. Hou, C. Jin, and W. Zhu, "AVQA: A dataset for audio-visual question answering on videos," in *Proceedings of the 30th ACM International Conference on Multimedia*, ser. Mm '22. Lisboa, Portugal and New York, NY, USA: Association for Computing Machinery, 2022, pp. 3480–3491.

[14] S. Lipping, P. Sudarsanam, K. Drossos, and T. Virtanen, "Clotho-AQA: A Crowdsourced Dataset for Audio Question Answering," in *30th European Signal Processing Conference, EUSIPCO 2022, Belgrade, Serbia, August 29 - Sept. 2, 2022*. IEEE, 2022, pp. 1140–1144.

[15] S. Ghosh, A. Seth, S. Kumar, U. Tyagi, C. K. R. Evuru, R. S, S. Singh *et al.*, "CompA: Addressing the Gap in Compositional Reasoning in Audio-Language Models," in *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria*, May 2024.

[16] P. Primus, F. Schmid, and G. Widmer, "TACOS: Temporally-aligned Audio CaptiOnS for Language-Audio Pretraining," May 2025.

[17] G. Wijngaard, E. Formisano, M. Esposito, and M. Dumontier, "AudSemThinker: Enhancing Audio-Language Models through Reasoning over Semantics of Sound," May 2025.

[18] Woods Hole Oceanographic Institution, "Watkins Marine Mammal Sound Database," https://whoicf2.whoi.edu/science/B/whalesounds/index.cfm.

[19] I. Trowitzsch, J. Taghia, Y. Kashef, and K. Obermayer, "The NIGENS General Sound Events Database," Feb. 2019.

[20] C. Marinoni, R. F. Gramaccioni, C. Chen, and D. Comminiello, "L3DAS23 - 2023 IEEE ICASSP Grand Challenge," https://www.l3das.com/icassp2023/.

[21] S. Adavanne, A. Politis, and T. Virtanen, "TAU spatial sound events 2019 - ambisonic and microphone array, development datasets," Feb. 2019.

[22] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events," in *Proc. IEEE ICASSP 2017*. New Orleans, LA: IEEE, 2017, pp. 776–780.

[23] X. Ju, Y. Gao, Z. Zhang, Z. Yuan, X. Wang, A. Zeng, Y. Xiong *et al.*, "Miradata: A large-scale video dataset with long durations and structured captions," *Advances in Neural Information Processing Systems*, vol. 37, pp. 48 955–48 970, 2024.

[24] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, ser. Icml '09. Montreal, Quebec, Canada and New York, NY, USA: Association for Computing Machinery, 2009, pp. 41–48.

[25] J. L. Elman, "Learning and development in neural networks: The importance of starting small." *Cognition*, vol. 48, no. 1, pp. 71–99, Jul. 1993.

[26] M. Abdin, J. Aneja, H. Behl, S. Bubeck, R. Eldan, S. Gunasekar, M. Harrison *et al.*, "Phi-4 Technical Report," Dec. 2024.

[27] F. Kreuk, G. Synnaeve, A. Polyak, U. Singer, A. Défossez, J. Copet, D. Parikh *et al.*, "AudioGen: Textually Guided Audio Generation," in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023.

[28] Z. Shao, P. Wang, Q. Zhu, R. Xu, J. Song, X. Bi, H. Zhang *et al.*, "DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models," Apr. 2024.

[29] A. Yang, B. Yang, B. Zhang, B. Hui, B. Zheng, B. Yu, C. Li *et al.*, "Qwen2.5 technical report," Jan. 2025.

[30] E. J. Hu, y. shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "LoRA: Low-rank adaptation of large language models," in *International Conference on Learning Representations*, 2022.

[31] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.

[32] B. T. Willard and R. Louf, "Efficient Guided Generation for Large Language Models," Aug. 2023.