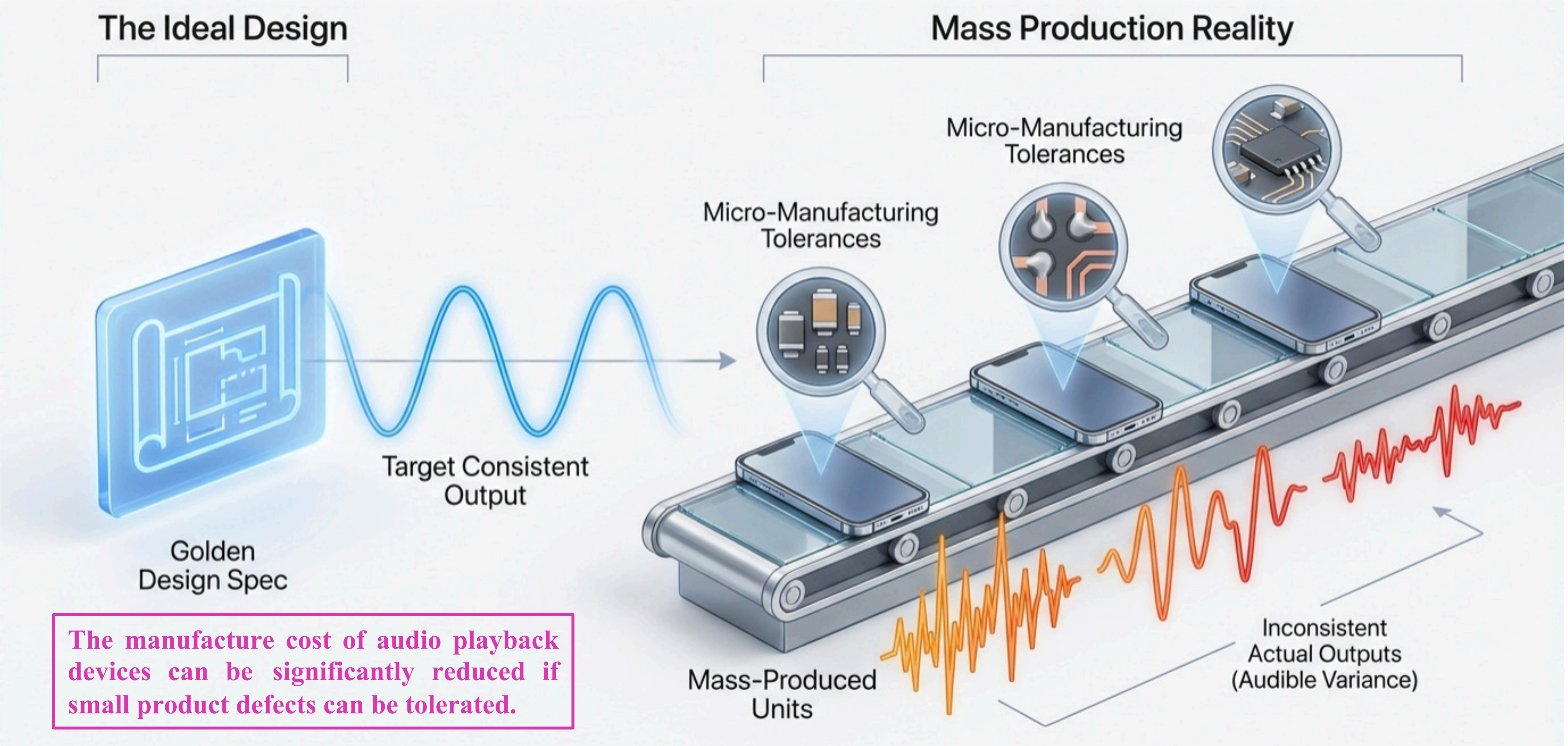
High Sample Rate Audio Generation Using Neural Large Language Models for Mitigating Device Manufacturing Defects

Baicheng Huang[1], Xinyi Pan[2], Haiwei Chai[2] (Oral Presenter), Feng Zhu[2], Dong Liu[2], Xiaoyong Pan[2]

1. *Tsinghua University, Beijing, China;*
2. *Conceptual Computing, Cambridge, MA, USA*

Introduction & Industrial Pain Point



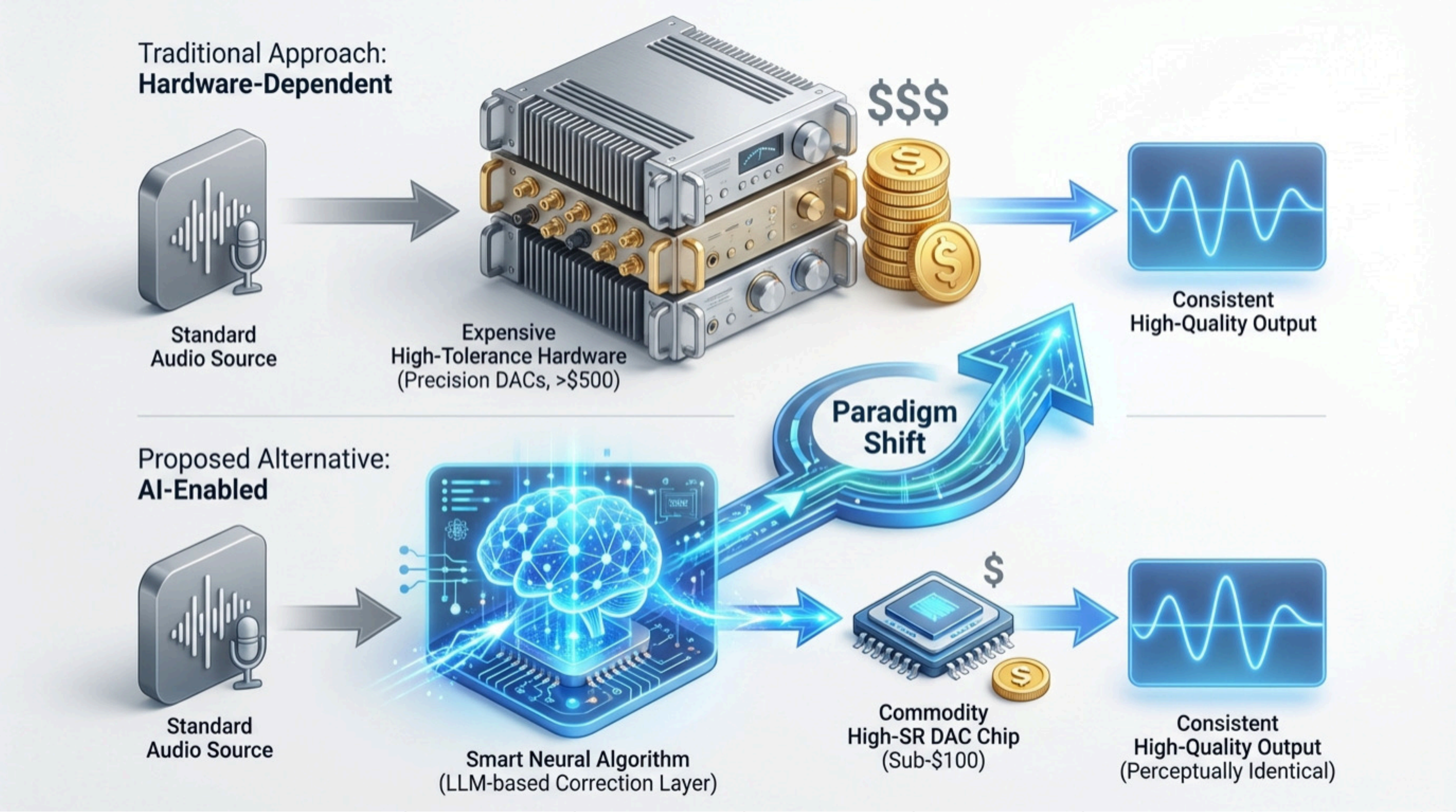
The manufacture cost of audio playback devices can be significantly reduced if small product defects can be tolerated.

Proposed Framework: AI-Defined Audio Quality

The Paradigm Shift: Moving away from relying solely on expensive, high-tolerance physical components to ensure consistency.

•**Smart Neural Layer**: Utilizing a specialized LLM-based algorithm to pre-process the audio, effectively "masking" potential hardware deficiencies before they occur.

•**Economically efficient hardware:** The algorithm is expected to be compatible with existing, low-cost high-sampling-rate digital-to-analog converter (DAC) chips **in the future**.



Concept: Audio Up-sampling for Masking Defects

•**Core Idea:** Using Large Language Models (LLMs) to up-sample audio to 384 kHz/32 bit or higher.

•**The Masking Effect:** The up-sampling algorithm is designed to inherit aural characteristics that mask subtle hardware deviations, making low-cost devices sound like high-end ones.

•**Software vs. Hardware:** This process is algorithm-based, making it significantly cheaper than conventional physical quality control.

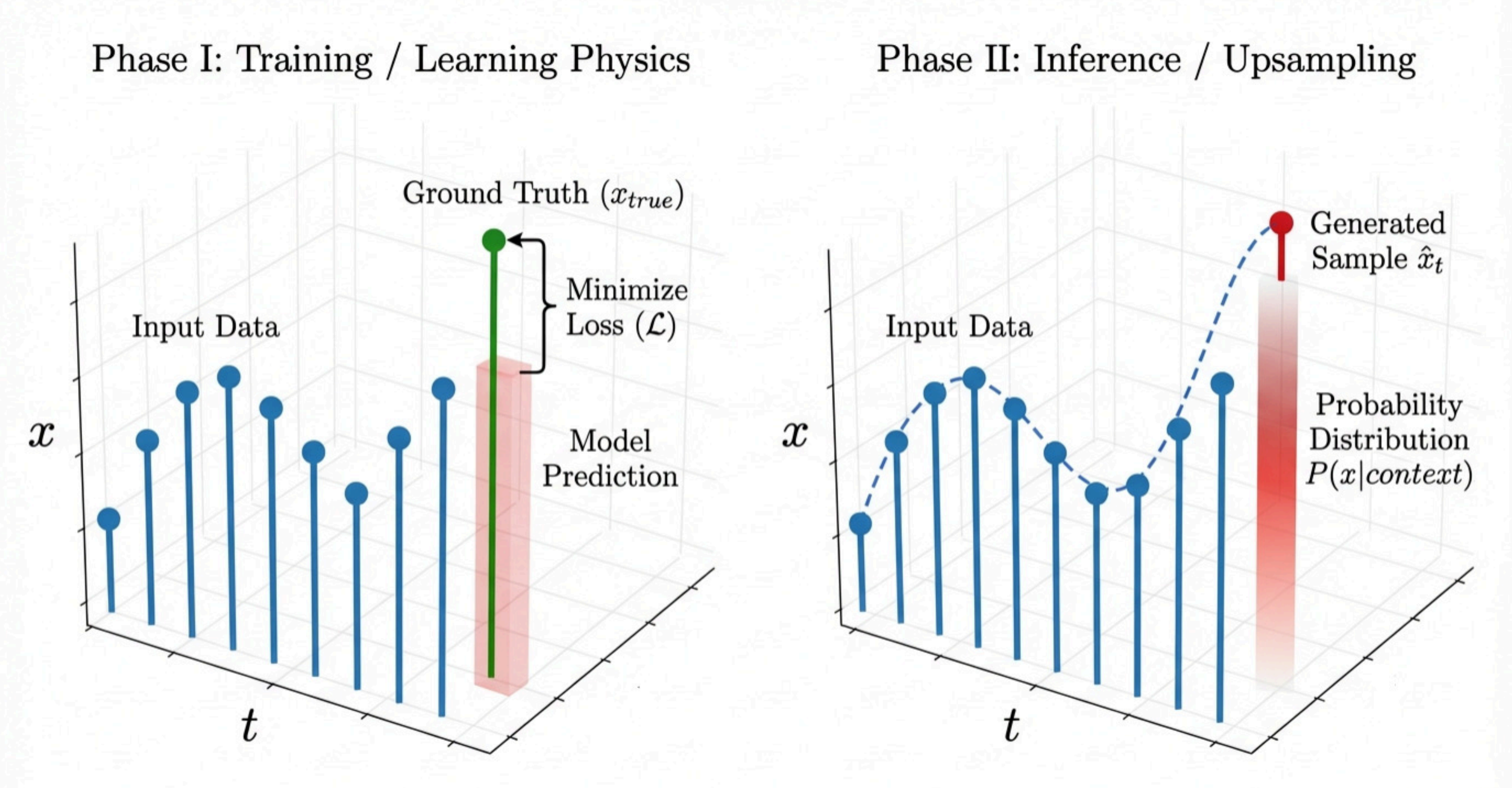


Fig. 1. Sequence prediction model for audio upsampling. (a) schematic illustration for model training; (b) illustration for inference.

Suppressing Artifacts: Branching Scheme

•**The Challenge:** Up-sampling can introduce high-frequency noise or annoying artifacts.

•**The Solution:** A multiple branching scheme for inference.

(a) Direct up-sampling/True sample.

(b) Predicted from a previous neural prediction.

(c) Predicted from a previous direct sample.

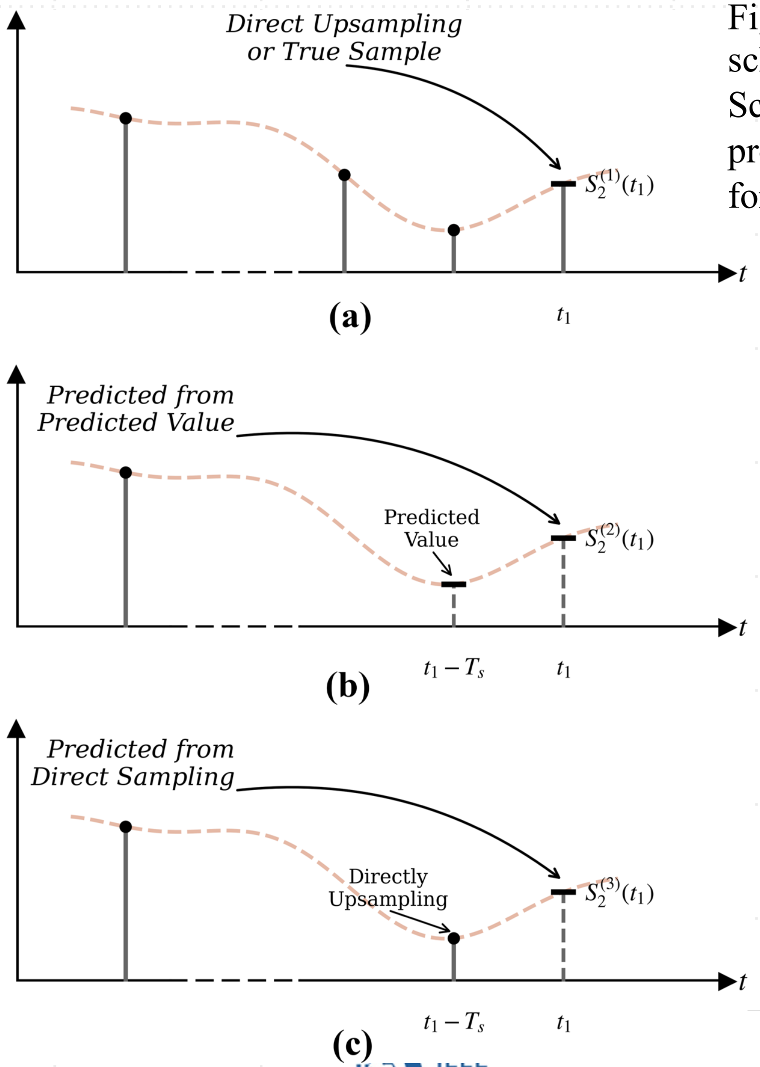


Fig. 2. Branching schemes for inference. Schemes (a), (b) and (c) provide three choices for next sample.

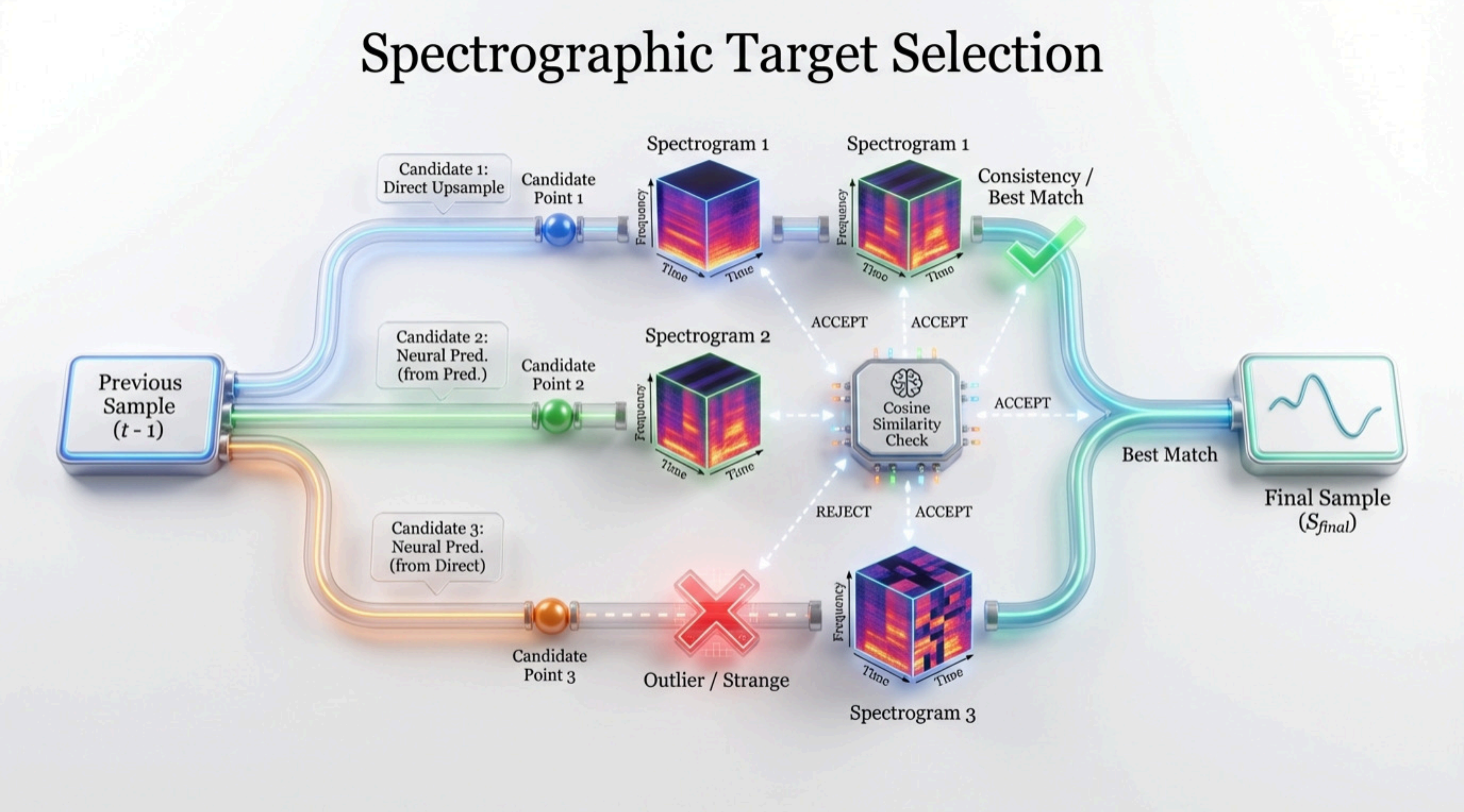
Spectrographic Target Smoothing

•**Comparison Logic:** The system generates spectrograms for the three branching choices and calculates **cosine similarity**.

•**Selection Process:** Discard the "strange" fragment (outlier).

•Select the fragment most similar to the previous step's choice to ensure temporal smoothness.

•**Benefit:** Preserves high-frequency information while removing up-sampling artifacts.



Detail: Audio Tokenization & Serialization

Hierarchical Quantization (DAC)

•**Multi-Codebook Structure**: Decomposes each audio frame into **9 parallel discrete codes**.

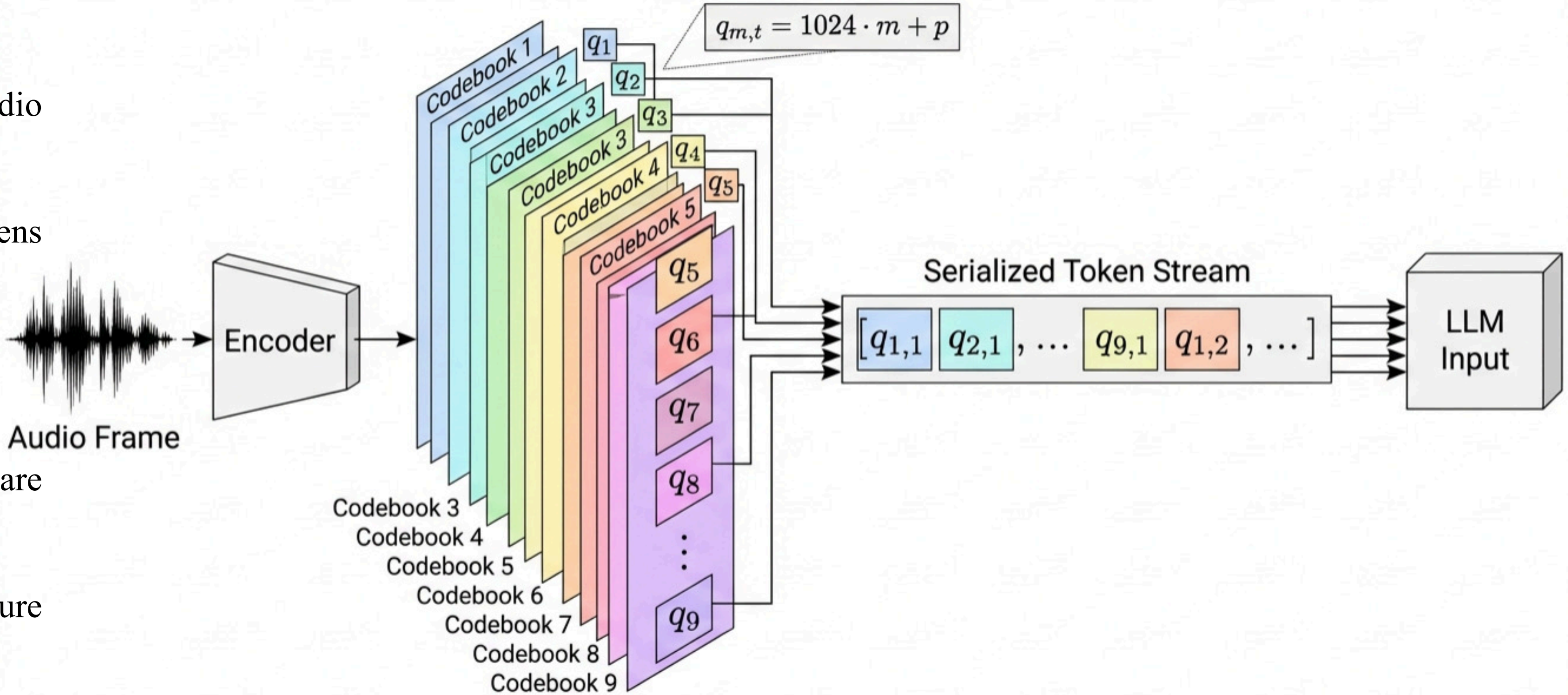
•**Vocabulary Mapping**: Maps 9 codebooks (1,024 tokens each) into a **9,216 non-overlapping integer space**.

1D Sequence Generation

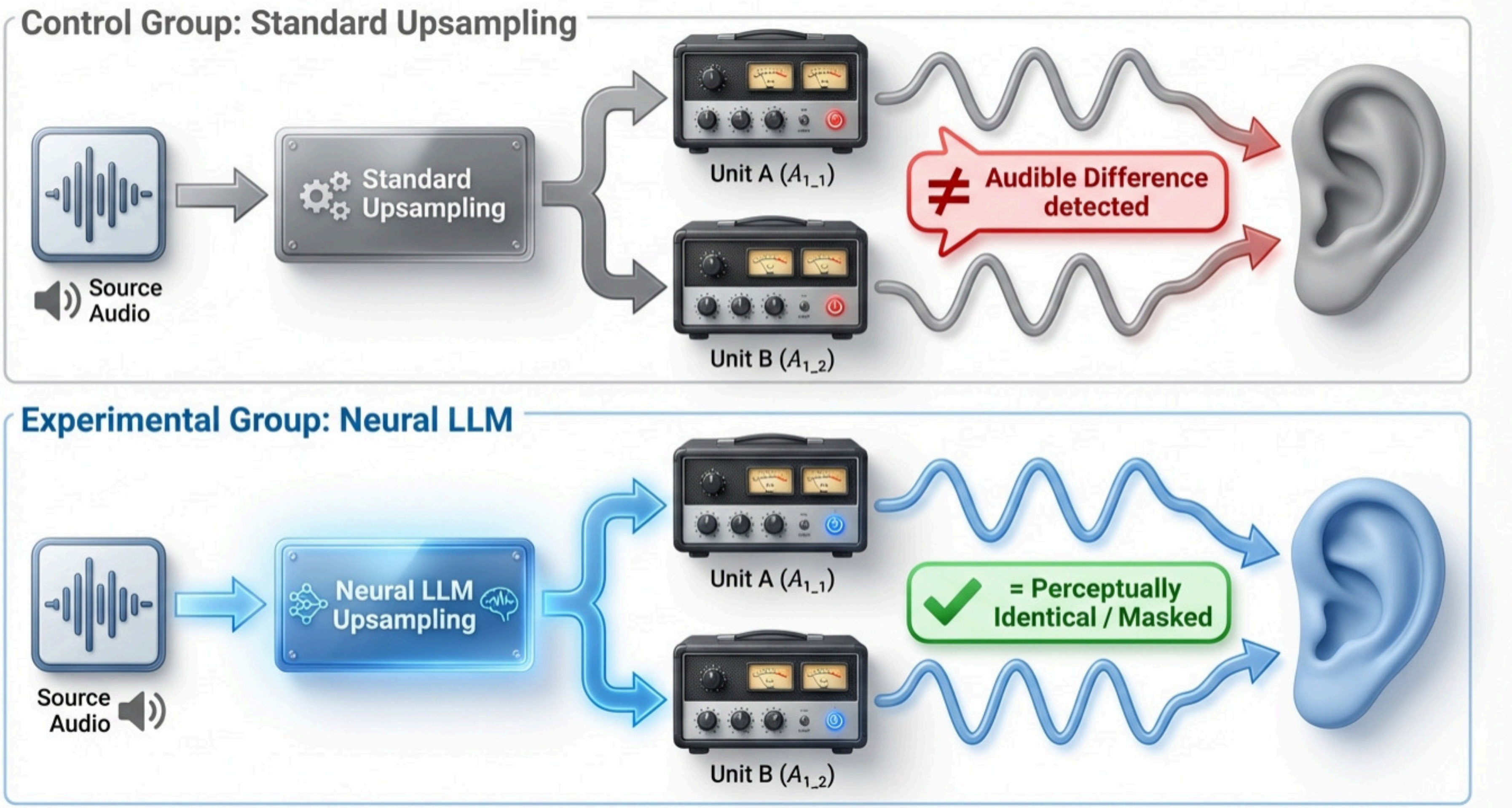
•**Interleaving**: Parallel codes from a single timestep are flattened sequentially.

•**Frame Alignment**: Audio is precisely padded to ensure length is a multiple of the model stride (512 samples).

•**Result**: A continuous "token stream" that serves as the final input for LLM training.

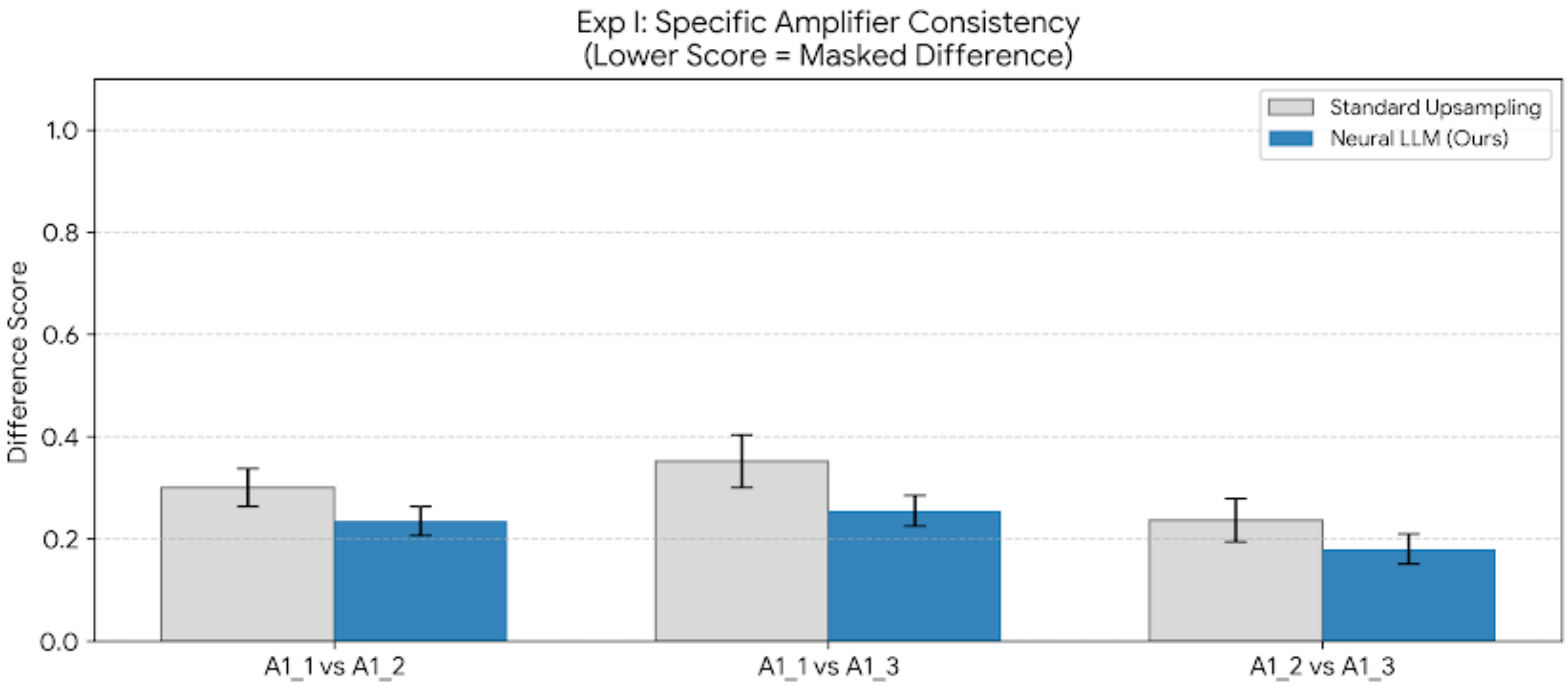


Experimental Setup



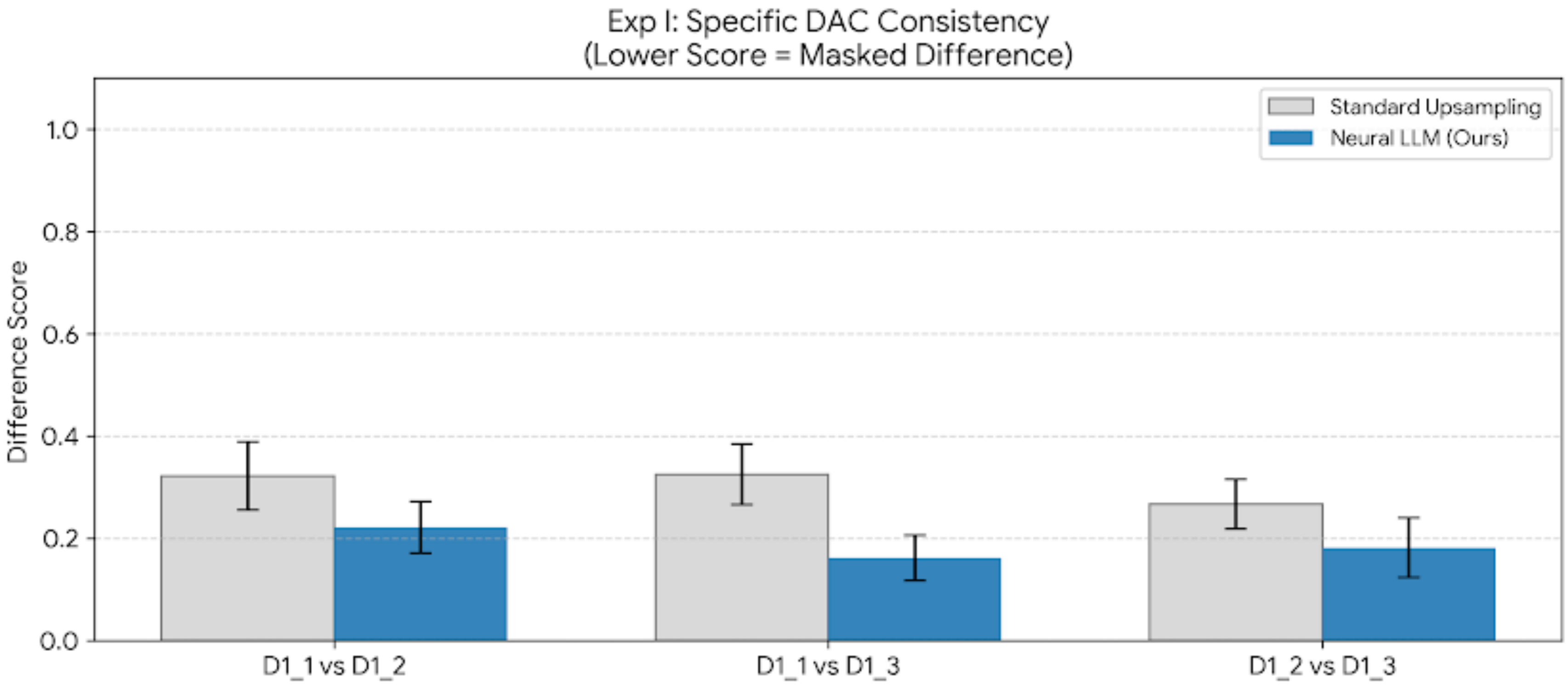
Experiment I: Device-Specific Masking

- **Setup:** Trained the LLM for one specific audio equipment model (using 3 identical units).
- **Protocol:** 12 experienced listeners (L1-L12) compared hardware labeled A1_1, A1_2 (Amplifiers) and D1_1, D1_2 (DACs).



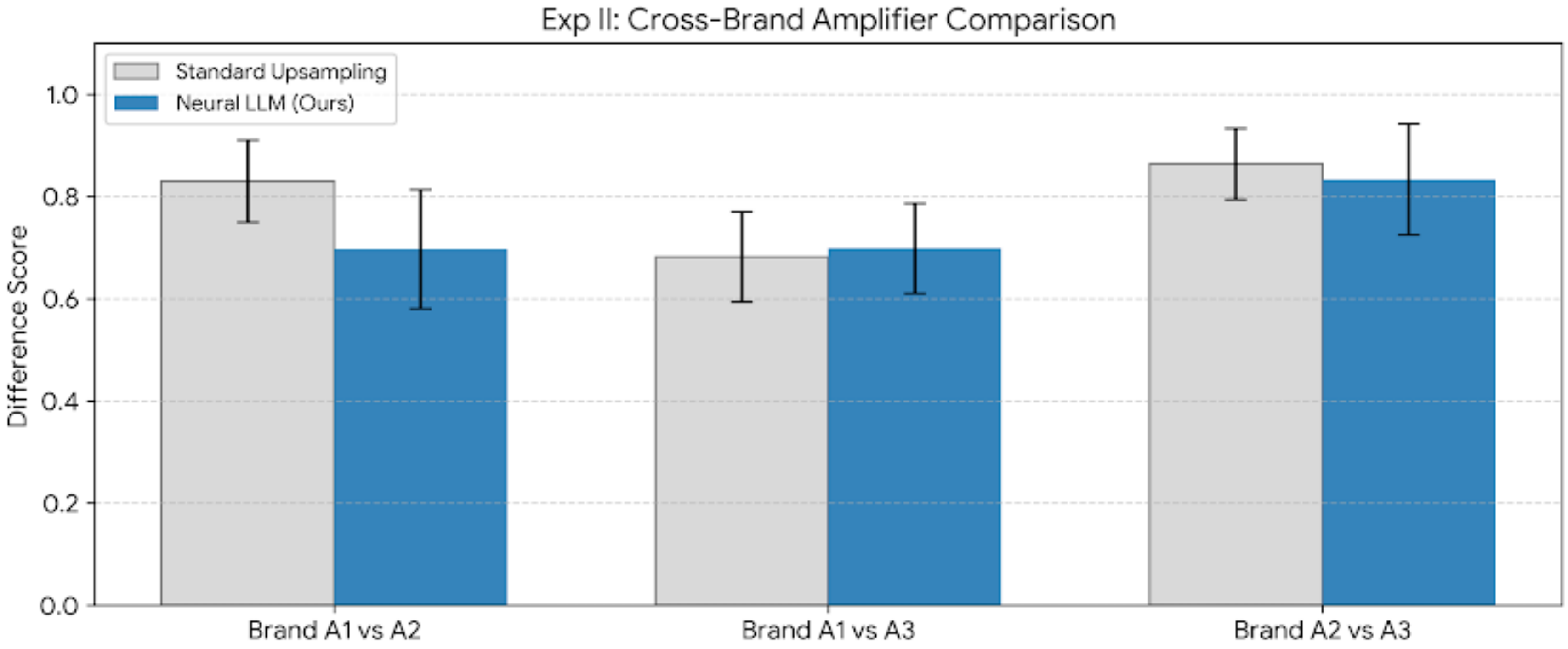
Experiment I: Device-Specific Masking

- **Setup:** Trained the LLM for one specific audio equipment model (using 3 identical units).
- **Protocol:** 12 experienced listeners (L1-L12) compared hardware labeled A1_1, A1_2 (Amplifiers) and D1_1, D1_2 (DACs).



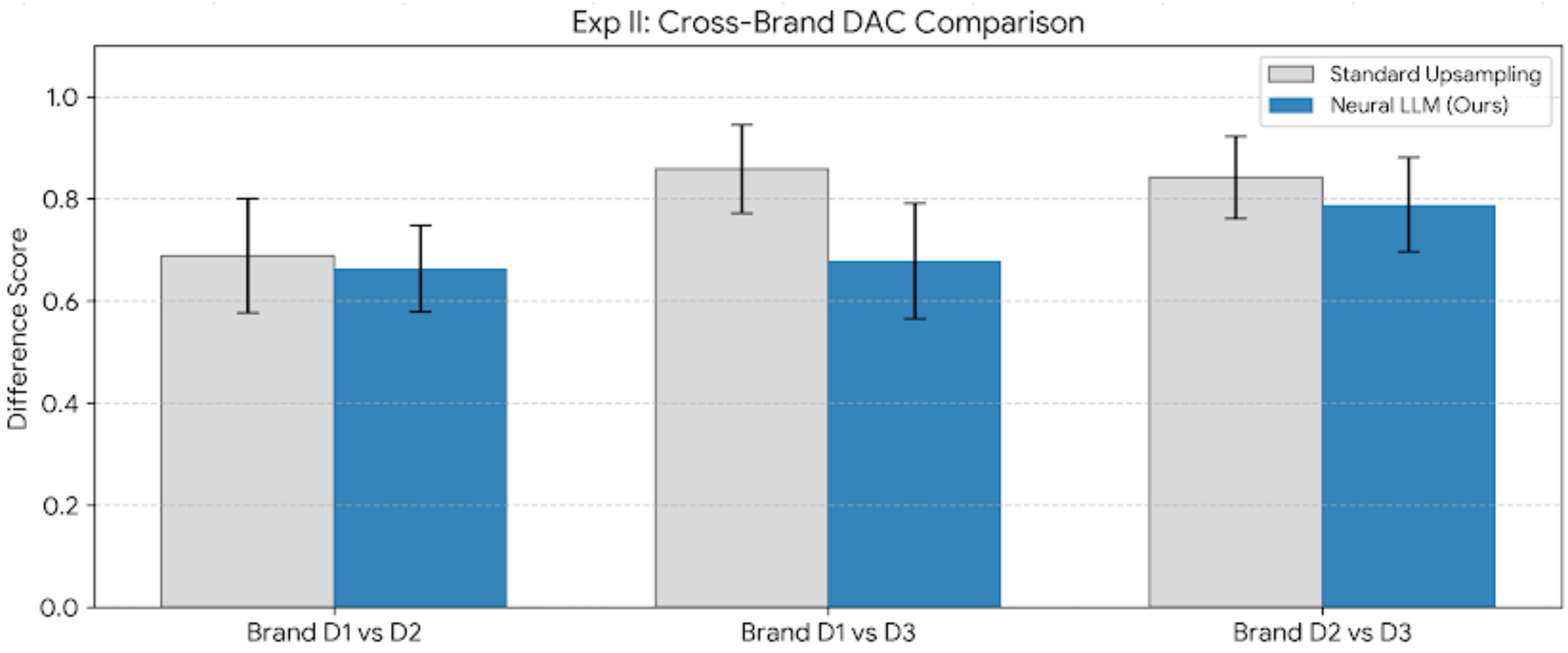
Experiment II: Cross-Brand Masking

•**Setup:** Tested the algorithm across devices of different brands (e.g., Amplifier A1 vs. A2 vs. A3).
•**Findings (Table III & IV):** Masking differences between totally different hardware architectures is significantly more difficult than masking deviations within the same model.



Experiment II: Cross-Brand Masking

•**Setup:** Tested the algorithm across devices of different brands (e.g., Amplifier A1 vs. A2 vs. A3).
•**Findings (Table III & IV):** Masking differences between totally different hardware architectures is significantly more difficult than masking deviations within the same model.



Conclusion & Economic Impact

- **Key Achievement:** Neural LLM-based up-sampling effectively masks manufacturing deviations in amplifiers and DACs.

- **Economic Value:** Useful for reducing prices of high-fidelity devices by substituting expensive electrical components with software and cheap DAC chips.

- **Future Work:** Extending the model to mask deviations in low-cost high-sampling-rate digital-to-analog converter chips or loudspeakers, headphones and earphones.

Thank You & Q&A

Thank you for your attention!

- **Contact Info: Haiwei Chai (chaihw@conceputing.com)**