Benchmarking Deep Learning vs. Statistical Signal Processing for Real-Time Sleep Apnea Detection

Researcher: Solomon Nderitu Wahome (Non-Trivial Fellow)

Abstract

Smartphone-based acoustic monitoring offers a scalable solution for diagnosing Obstructive Sleep Apnea (OSA) but faces significant challenges in uncontrolled bedroom environments where environmental noise masks diagnostic cues. Current industry approaches prioritize large-scale Generative AI models, which are often computationally unviable for mobile edge devices. This research investigated the "Efficiency-Accuracy Trade-off" by benchmarking five noise-suppression architectures—Generative AI (Demucs), Lightweight AI (DeepFilterNet), and Classical DSP (Spectral Gating, Wiener, Wavelet)—on simulated mobile hardware (CPU-only execution). A synthetic stress-test dataset was engineered by mixing biomedical recordings with categorized noise at 0dB and -5dB SNRs. Results revealed an "Efficiency Paradox": while Demucs provided strong semantic separation, it averaged ~4.0s inference latency, exceeding real-time thresholds by two orders of magnitude. Conversely, traditional Spectral Gating (~20ms latency) demonstrated superior signal fidelity for stationary noise profiles. The study concludes that massive AI models are inefficient for continuous monitoring and proposes "AdaptiveSleep," a cascaded hybrid architecture utilizing DSP as a low-power sentinel.

Methodological Challenge: Overcoming Circular Bias

Initial experimental design utilized a custom Classifier to judge audio quality. However, analysis revealed this introduced **circular bias**, as the model overfitted to the specific noise patterns in the training data rather than measuring true signal restoration.

To resolve this, the evaluation protocol was completely re-engineered to utilize **YAMNet (Transfer Learning)**. By extracting embeddings from an unrelated dataset (AudioSet) and calculating Cosine Similarity, I established an objective, third-party standard for measuring semantic signal preservation.

Key Findings

- **The Efficiency Paradox:** Generative AI (Demucs) achieved high semantic separation but averaged **~4.0 seconds** latency, disqualifying it for real-time mobile use.
- **The Hybrid Solution:** Classical Spectral Gating (~20ms latency) proved superior for stationary noise (fans), while AI excelled at transient noise (sirens).

Data Visualizations

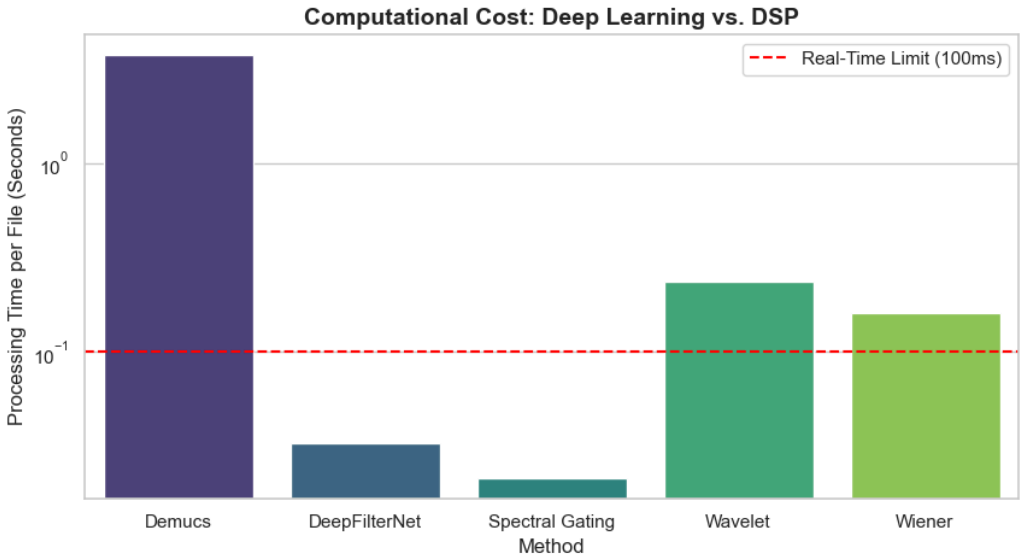

Figure 1: Inference latency comparison on CPU. Note the logarithmic gap between Demucs (AI) and Spectral Gating (DSP).

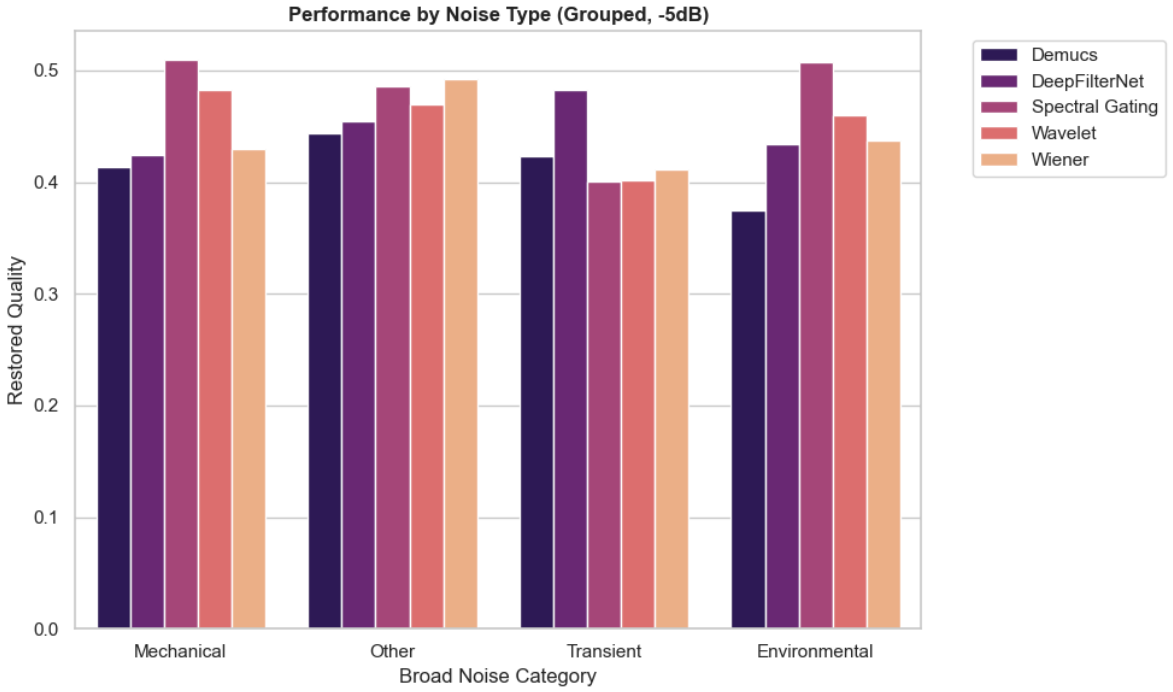

Figure 2: Restoration quality (Cosine Similarity) by noise type. Note that simple DSP outperforms AI on common stationary noise.

Proposed Architecture

Based on these findings, I propose **"AdaptiveSleep,"** a cascaded architecture that uses ultra-low-power DSP as a default sentinel and only wakes the AI model when high-complexity transient noise is detected.