

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение
высшего образования

«Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Лабораторная работа

по дисциплине

Разработка мультимедийных приложений

на тему

«Восстановление звука через OpenVINO-AI

(Лаб. 5)»

Выполнили студенты:

Владимиров Владислав Александрович Р3322

Преподаватель:

Андреев Артём Станиславович

г. Санкт-Петербург

2026

1. Введение

Цель работы — оценка эффективности метода восстановления звуковых сигналов с использованием нейросетевой модели DeepFilterNet3, интегрированной в среду Audacity через OpenVINO. В качестве исходного материала использована запись голоса (VoiceA), к которой был добавлен шум (SNR 1 dB). Восстановление выполнялось алгоритмом DeepFilterNet3.

2. Метрики оценки качества

Для количественной оценки использовались следующие метрики:

- SNR (Signal-to-Noise Ratio) — отношение мощности сигнала к мощности шума;
- Segmental SNR — усреднённый SNR по сегментам 30 ms с клиппированием $[-10, 35]$ dB;
- LSD (Log-Spectral Distance) — среднеквадратичное логарифмическое спектральное расстояние;
- Коэффициент корреляции Пирсона между сигналами;
- RMSE (Root Mean Square Error) — среднеквадратичная ошибка;
- Фактор шумоподавления — отношение дисперсии шума до и после обработки.

3. Результаты

3.1. Сводная таблица метрик

Таблица 1 – Сравнение метрик для зашумлённого и восстановленного сигналов

Метрика	Шумный	Восстановленный	Улучшение
SNR (dB)	0.92	2.36	+1.44
Segmental SNR (dB)	−1.45	1.26	+2.71
LSD (dB)	73.78	31.34	−42.44
Корреляция	0.6664	0.6633	−0.0031
RMSE	9202	7801	15.2%
Шумоподавление (dB)	—	—	+1.44

3.2. Анализ результатов

DeepFilterNet3 демонстрирует следующие результаты:

- SNR +1.44 dB: прирост отношения сигнал/шум положительный, однако скромный. Исходный SNR 0.9 dB был крайне низким — модель работает в условиях сильного шума.

- Segmental SNR +2.71 dB: посегментный анализ показывает более существенное улучшение на коротких интервалах, что говорит о качественном подавлении шума в паузах.
- LSD –42.44 dB: резкое снижение логарифмического спектрального расстояния — спектральная форма восстановленного сигнала значительно ближе к оригиналу, чем у шумного.
- Корреляция: практически не изменилась (0.666 против 0.663). Это связано с тем, что шум аддитивный и не разрушает линейную структуру сигнала, а слабое подавление не восстанавливает корреляцию.
- RMSE: снизилась на 15% — амплитудная ошибка уменьшилась.

Ключевое наблюдение: модель эффективно подавляет шум в спектральной области (LSD падает в 2.4 раза), но вклад шума в амплитудную ошибку остаётся значительным. Это характерно для DeepFilterNet, которая работает в частотной области.

3.3. Визуализация

На рисунке 1 представлен комплексный анализ сигналов.

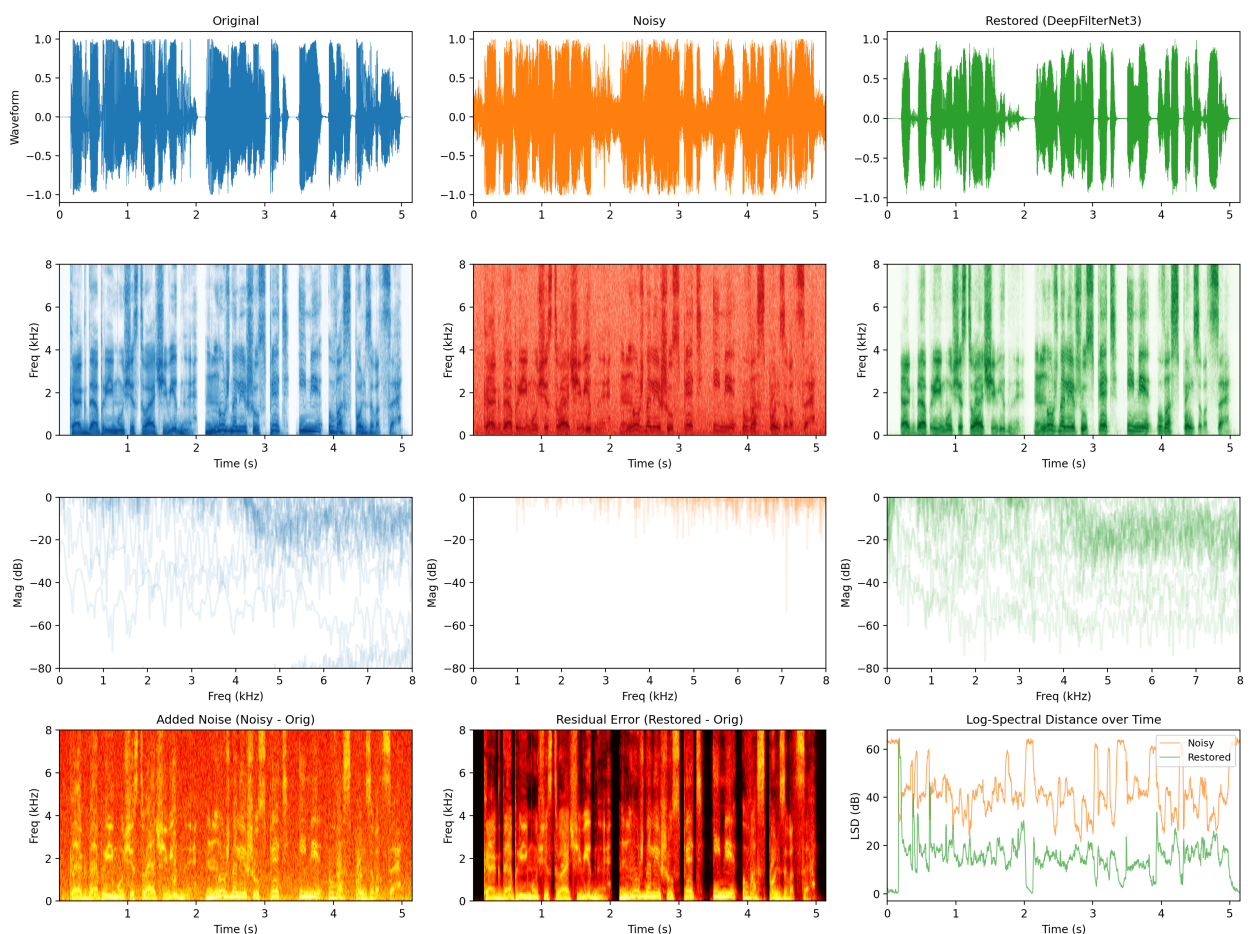


Рисунок 1 – Сравнение исходного, зашумлённого и восстановленного сигналов. Строки: (1) осциллограммы, (2) спектрограммы, (3) усреднённые спектры, (4) спектрограмма шума, ошибки восстановления и LSD(t).

3.4. Спектральный анализ

По спектрограммам (рис. 1, строка 2) видно:

- Шум занимает широкий частотный диапазон (0–8 kHz) и маскирует формантные структуры голоса.
- Восстановленный сигнал имеет более чистую спектральную картину; шумовой фон в паузах подавлен.
- Высокочастотные компоненты (>4 kHz) восстанавливаются хуже — часть гармоник теряется.

Спектрограммы остаточной ошибки (строка 4, центр) показывают, что наибольшая ошибка восстановления сосредоточена в области 1–3 kHz, где энергия голоса максимальна — модель осторожничает и не подавляет шум в этой области, чтобы не исказить речь.

4. Заключение

DeepFilterNet3 в Audacity (OpenVINO) показал следующие результаты:

- Улучшение SNR на +1.44 dB и Segmental SNR на +2.71 dB;
- Значительное снижение спектрального искажения (LSD с 73.8 до 31.3 dB);
- Снижение RMSE на 15%;
- Корреляция с оригиналом практически не изменилась.

Метод эффективен для подавления широкополосного шума, особенно заметен эффект в паузах между речевыми сегментами. При очень низком исходном SNR (~ 1 dB) полное восстановление невозможно, однако модель демонстрирует разумный компромисс между подавлением шума и сохранением речевого сигнала. Для лучших результатов рекомендовано использовать более высокое качество входного сигнала (SNR 5–10 dB) или комбинировать DeepFilterNet с предварительной фильтрацией.