

Шаг 1 - какой тип метрики?

- Эталон

Есть чистый оригинал? Полноэталонная метрика (PESQ, POLQA, ViSQOL, PEAQ). Лучше всего в сборке и лаборатории.
- Без эт.
- Эталон-человек

Нет оригинала (живой трафик)? Метрика без эталона (DNSMOS P.835, NISQA). Построены на ML.
- Все они предсказывают Mean Opinion Score (MOS). Панель живых слушателей остаётся истиной.

Шаг 2 - выберите полноэталонную метрику

Метрика	Для чего	Диапазон	Статус 2026	Цена
PESQ (P.862)	речь / телефония	до 7 кГц	удалён 01.2024	бесплатно
POLQA (P.863)	речь / телефония	до 20 кГц	актуален	по лицензии
ViSQOL v3	речь + музыка	16 / 48 кГц	поддержив.	беспл. (Apache 2.0)
PEAQ (BS.1387)	муз. кодеки	полнополосн.	актуален (2023)	стандарт

Факты, на которых спотыкаются

- PESQ отозван в 2018 и удалён из каталога ITU 05.01.2024 - но в 2024 его использовали ~4600 статей (бесплатен; POLQA - по лицензии).
- POLQA закрывает два пробела PESQ: достаёт до полнополосного аудио и корректно оценивает time-warping (изменения скорости от кодека/джиттера).
- ViSQOL превращает аудио в спектрограмму и сравнивает картинки через NSIM (мера сходства изображений). Режим 48 кГц - редкий бесплатный вариант для музыки.
- PEAQ выдаёт Objective Difference Grade (ODG), от 0 (незаметно) до -4 (очень раздражает), для кодеков AAC/MP3.
- «AAC-Q» - не стандарт; это качество муз. кодеков, измеряемое PEAQ / ViSQOLAudio / вендорскими инструментами.

Метрика в CI - дисциплина

- ☐ Подбирайте метрику под контент - речевая метрика на музыке даёт уверенную чепуху.
- ☐ Зафиксируйте набор из 30-50 клипов, который модель не видит; храните оценки по клипам, не только среднее.
- ☐ Порог в CI - из статистики: падать только при разнице больше стандартной ошибки ($s / \sqrt{N}$).
- ☐ Никогда не обучайте модель на метрике - закон Гудхарта делает оценку обманываемой.
- ☐ Подкрепляйте периодическим слуховым тестом по P.808; при расхождении метрики и людей правы люди.