

Детектор речи, прерывистая передача, комфортный шум, настройки Opus и два режима отказа.

VAD: четыре поколения

Порог по энергии	Громко = речь	Ломается в шумной комнате
Классический (G.729-B)	4 признака + слежение за шумом	Эталон телекома, 1996
WebRTC VAD	6 полос + смесь гауссиан	Агрессивность 0–3; фреймы 10/20/30 мс
Silero VAD (нейро)	~309K параметров, ~1–2 МБ, v5 2024	Куски 32 мс; многоязычность; точнее

Два режима отказа (+ фикс)

Ложноотрицательная	Речь принята за тишину	Обрезает слова — это ненавидят
Ложноположительная	Шум принят за речь	Тратит трафик (только облачный счёт)
Hangover (фикс)	Хвост передачи после речи	Выкупает почти все обрезанные слова

Opus DTX — настройки, которые важны

usedtx по умолч. = 0 (выкл)	Нужно явно задать usedtx=1
Тишина: ~1 фрейм / 400 мс	против 1 фрейма / 20 мс при речи
Сброс ТОЛЬКО целых фреймов	метки RTP кратны 120
DTX против потери	приёмник читает пробел меток vs seq
НЕ добавлять CN из RFC 3389	Opus делает свой; сочетать не реком.
Речевой режим, не музыка	DTX включается в тракте SILK/речь

Арифметика трафика

Речь: ~60 байт нагрузки Opus каждые 20 мс = $60 \times 50 = 3\,000$ байт/сек.
 Простой (DTX): ~3 байта описателя каждые 400 мс = $3 \times 2,5 = 7,5$ байт/сек.
 Падение нагрузки на неактивном потоке ~ 99% (меньше с заголовками RTP/UDP/IP).
 Экономия по звонку — «десятки процентов»: зависит от доли речи и заголовков.

DTX включать или нет?

DTX ВКЛ: разговор — встречи, телемедицина, контакт-центр, обучение (микрофоны в простое).
 DTX ВЫКЛ: непрерывный звук — музыка, диджей-сеты, инсталляции, радионяни (тишина = сигнал).
 Мягкий VAD + щедрый hangover: клиника / тихий голос, где обрезка слов недопустима.
 Агрессивный DTX на микрофонах аудитории: большие вебинары, чтобы не топить медиасервер.

Запомнить

- ☐ VAD решает «речь / тишина» на фрейм; DTX действует, останавливая полную передачу во время тишины.
- ☐ Во время тишины слать крошечный описатель SID (уровень + спектр), не ничего — приёмник проигрывает комфортный шум.
- ☐ Комфортный шум не несёт звука, только рецепт: байт уровня в dBov плюс необязательные спектральные коэффициенты (RFC 3389).
- ☐ Тот же VAD питает эндоинтинг транскрипции, удаление мёртвого эфира и индикатор активного спикера.
- ☐ Настраивайте DTX вместе с шумоподавлением и усилением — в тракте захвата они не независимы.