

Как устроена оценка ЗВОНКОВ

Для разработчиков. Состояние на 24.09.2026. Подробности — в README репозитория `Buharestskie/amocrm_transcribe_service` и `Buharestskie/yandex_transcribe_service`.

Коротко

Все дашборды — МПП2, МОС и МПП1 — расшифровывают и оценивают звонки через один сервис: **amocrm_transcribe_service** (в конфигах дашбордов он называется «S2T apiscrm»). Он отдаёт запись в распознавание речи A5X, раскладывает реплики на менеджера и клиента и оценивает разговор языковой моделью A5X по промπτ, который прислал дашборд. Критерии оценки живут не в сервисе, а в «Настройках» каждого дашборда.

Путь звонка

1. **Звонок попадает в дашборд** из телефонии — Sipuni, МегаФон или своя PBX: вебхуком или синхронизацией истории.
2. **Дашборд решает, что оценивать**, по своим правилам: МПП2 — серия звонков после входа сделки в «Принимают решение», МОС — звонки по касаниям, рискам и рекомендациям, МПП1 — звонки охотников. Звонки одного клиента за день склеиваются в серию.
3. **Запрос в сервис** — `POST /internal/transcribe-analyze` на внутренний порт сервиса на `global` (адрес — в README): ссылки на записи, тип телефонии у каждой, промπτ и адрес для ответа. Сервис сразу отвечает `202`.
4. **Распознавание** — запись целиком уходит в `s2t.a5x.pro`. Левый канал становится `SPEAKER_00`, правый — `SPEAKER_01`.
5. **Роли** — кто из них менеджер, сервис решает по типу телефонии, не по тексту (таблица ниже).
6. **Оценка** — сегменты уходят в `llm.a5x.pro` (модель `qwen36-35b-a3b`): промπτ дашборда — системным сообщением, разговор — в виде `[секунда роль] текст`.
7. **Ответ** — сервис присылает дашборду обратный вызов: расшифровку, ответ модели (сырой и разобранный JSON) и расход — секунды аудио и токены. Дашборд сохраняет разбор и показывает его: МПП2 — на вкладке «Качество» и в виджете «Оценка разговоров» в карточке сделки, МОС — в касаниях, рисках и

рекомендациях, МПП1 — в карточке звонка.

Дашборд	Куда приходит ответ	Где промпты
МПП2	<code>/api/transcriptions/callback</code>	Настройки: две консультации, касания, тёплый пул
МОС	<code>/api/telephony/webhooks/transcribe</code>	Настройки: касания, риски, рекомендации
МПП1	<code>/api/v1/internal/transcribe/callback</code>	Настройки: гейт и QA-аудитор

Каналы по телефониям

Телефония	Менеджер	Клиент
Sipuni	правый канал	левый
МегаФон	левый	правый
Своя PBX	правый	левый

Запись должна быть стерео: моно-запись получает ошибку, роли не угадываются. Подключаете новую телефонию — сначала строка в этой таблице (`src/speaker_mapping.py`), иначе менеджер и клиент перепутаются во всех разборах.

Резерв: Yandex

yandex_transcribe_service (тоже на `global`) — распознавание SpeechKit и оценка YandexGPT в два этапа (гейт, затем полный разбор). Раньше был основным, с 17.09.2026 рабочих запросов нет: остался кнопкой ручной перерасшифровки «через Yandex» в МПП2 и МПП1. Каталог Yandex Cloud при этом живой и платный.

Ещё два репозитория — `assembly_transcription` (AssemblyAI) и `gemini_service` (Gemini) — нигде не развёрнуты.

Если оценки не приходят

- 1. **Жив ли сервис:** с `global` — `curl -s localhost:8083/health`.
- 2. **Доходят ли запросы:** `docker logs --since 1h amocrm_transcribe_service-apiscrm-transcribe-1 | grep transcribe-analyze`.

3. **Ошибки** — в GlitchTip, проект сервиса, и в обратных вызовах с полем ошибки: их дашборд показывает у звонка.
4. **Сервис перезапускали?** Задачи в работе живут в памяти и при перезапуске теряются. МОС повторяет звонок без ответа через час (потом через 30 минут и 2 часа, всего четыре попытки), МПП1 — задачей раз в 20 минут.
5. **Ответ пришёл, а разбора нет** — модель ответила не JSON: смотреть сырой `analysis_text` в данных звонка и промпт в «Настройках».

Что стоит денег

Каждый разбор — минуты распознавания A5X и токены модели. Ключ у распознавания и модели сейчас один: отзовут — встанут оба этапа. Правка промпта в «Настройках» действует только на новые звонки; старые заново не оцениваются.

Версия #2

Admin создал 2026-09-24 12:27:26 UTC

Admin обновил 2026-09-24 14:41:32 UTC