

На шаг ближе к метавселенной, или Как мы оживили аватары пользователей в VK Звонках

Алексей Шпагин — руководитель команды бэкенда VK Звонков, ВКонтакте.



Обо мне

 Руководитель команды бэкенда
VK Звонков

 10 лет работы в VoIP-телефонии
и видеозвонках

 Бэкграунд — разработчик C++

 В руководстве командами 5 лет



VK ЗВОНКИ

Бесплатные звонки без ограничений по времени и количеству участников. Для любых устройств.



Для учёбы и работы

Демонстрация экрана в разрешении 4K, трансляция, планирование и запись звонков.



Удобное управление звонками

Зал ожидания, управление микрофонами, функция «Поднять руку» и другие возможности модерации.

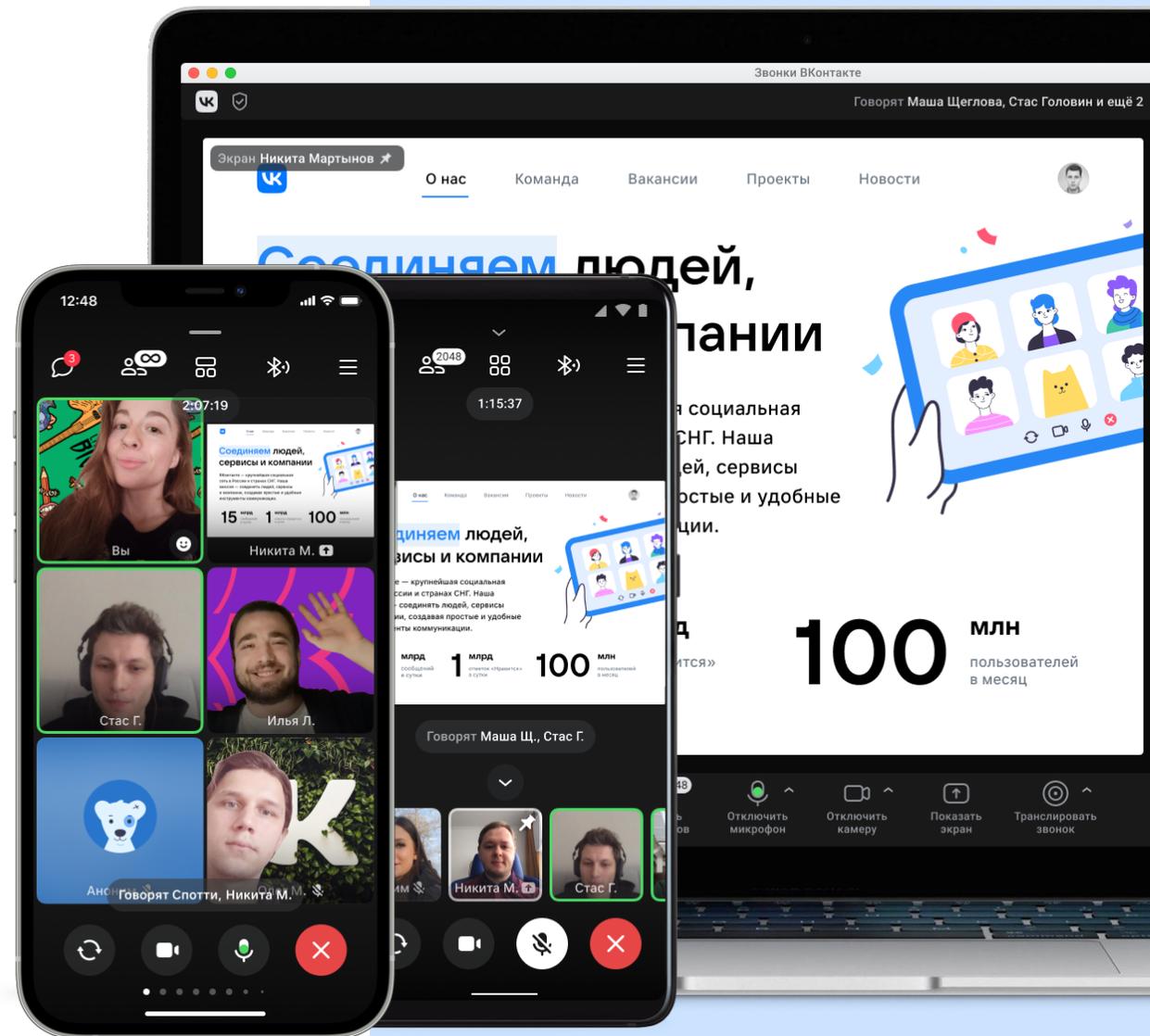


Технологичность

Продвинутое интеллектуальное шумоподавление, собственная AR-технология замены фона.

20 млн

пользователей общаются в VK Звонках ежемесячно





VK ЗВОНКИ

6 МЛН

ЗВОНКОВ В ДЕНЬ

20 МЛН

ПОЛЬЗОВАТЕЛЕЙ В МЕСЯЦ

15 ТЫС.

ОДНОВРЕМЕННЫХ ЗВОНКОВ

Содержание

1

Что такое анимированные аватары в Звонках

2

Как реализована анимация

3

Как происходит передача данных между пользователями

4

Обо что мы «споткнулись» при разработке

Что такое анимированные аватары в Звонках?





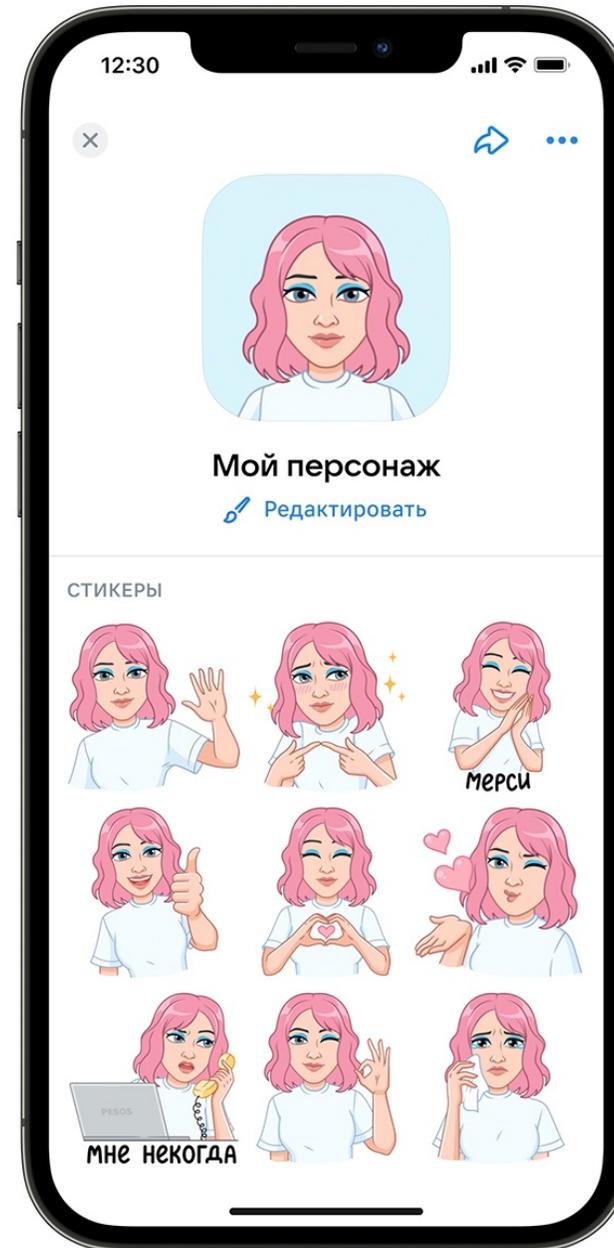
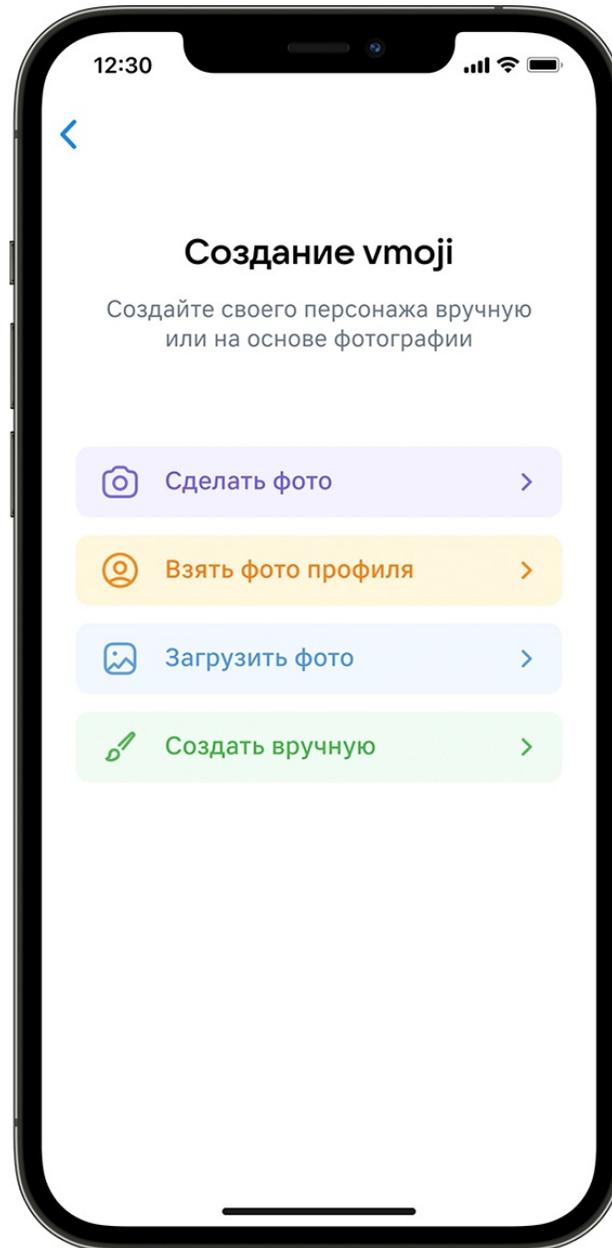
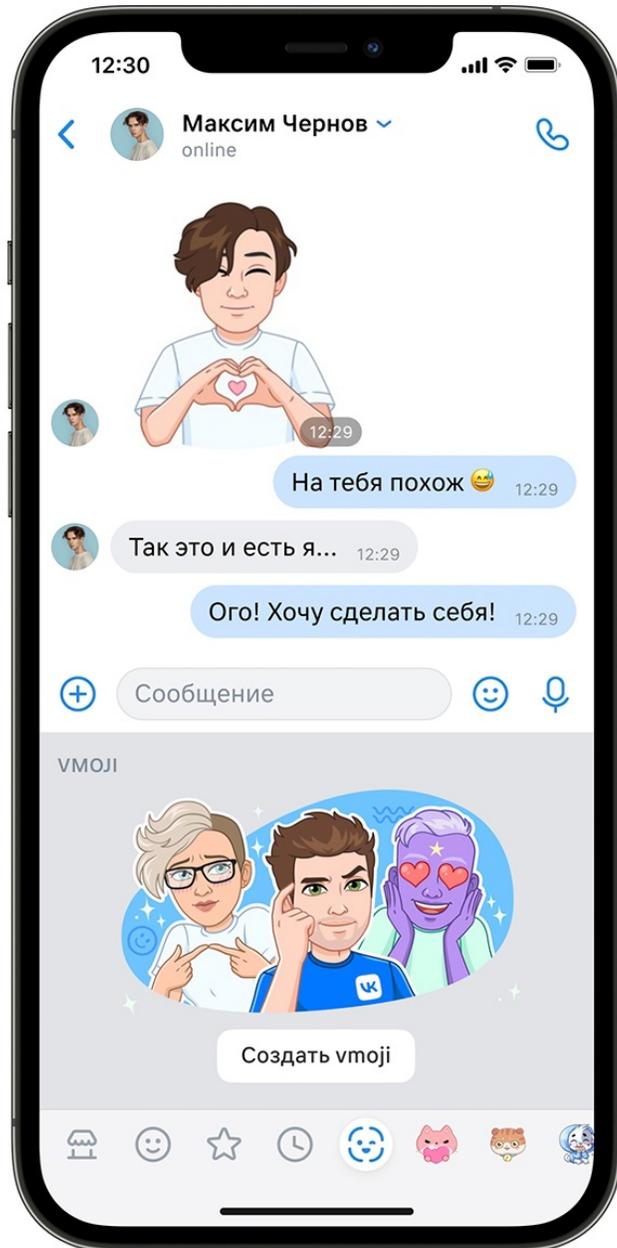
Создайте свой уникальный набор стикеров с vmoji

Стикеры с vmoji

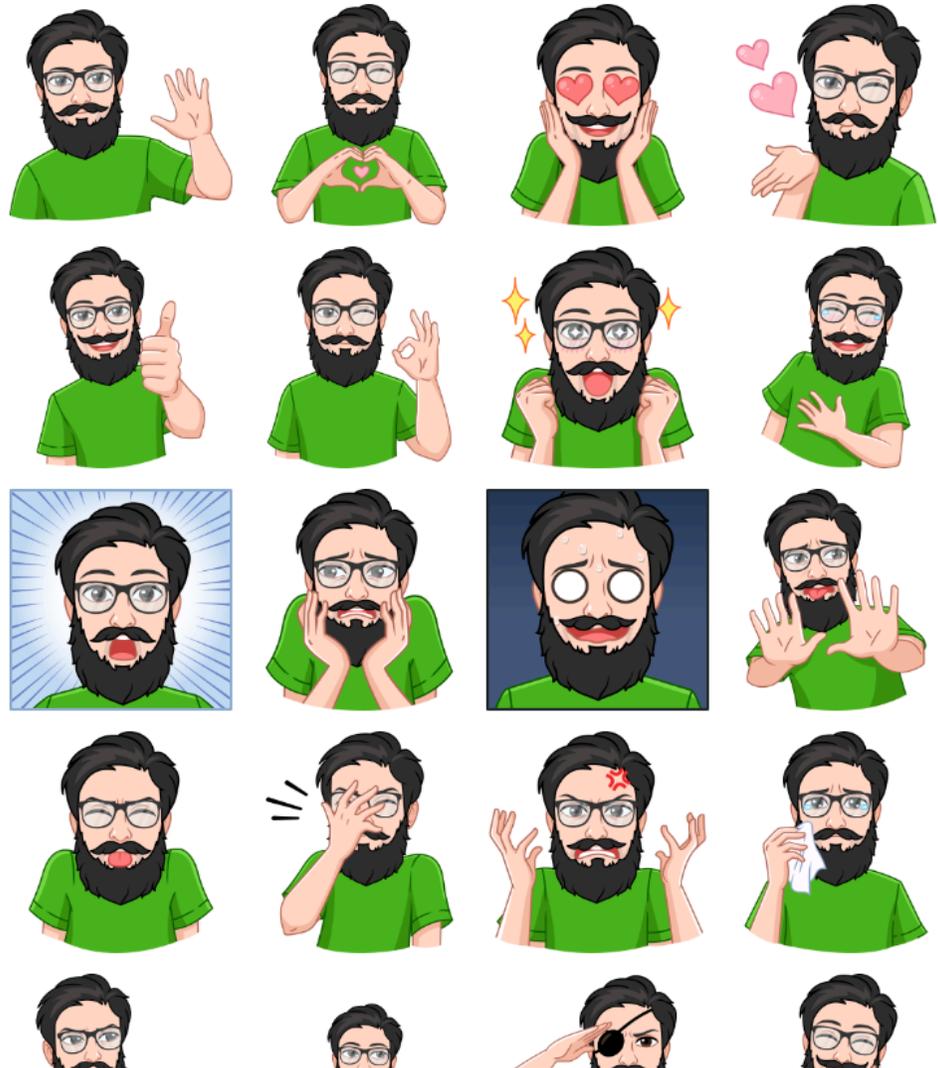
- Создаются автоматически с помощью ML на основе фото
- Доступно редактирование вручную
- В любой момент можно сгенерировать нового персонажа



Vmoji
пресс-релиз

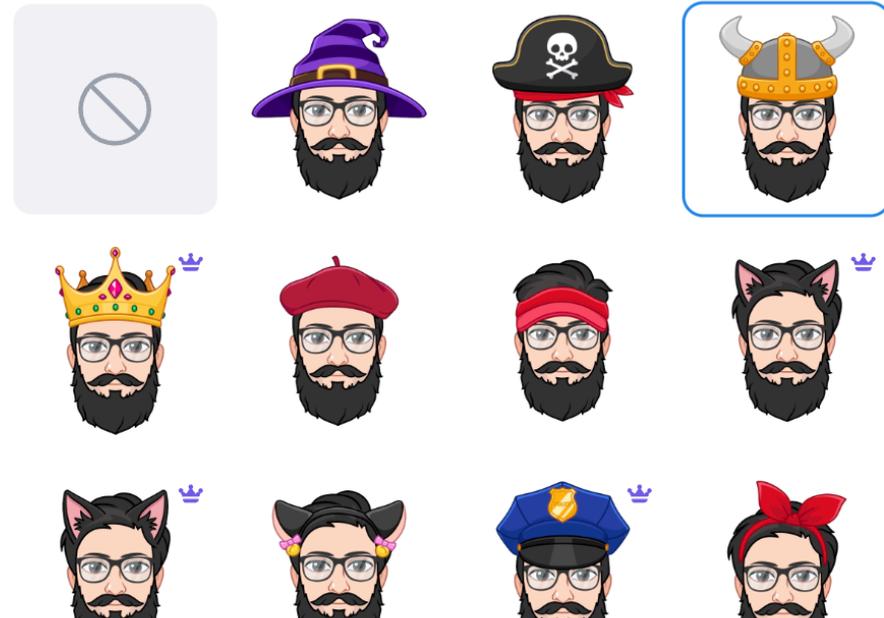


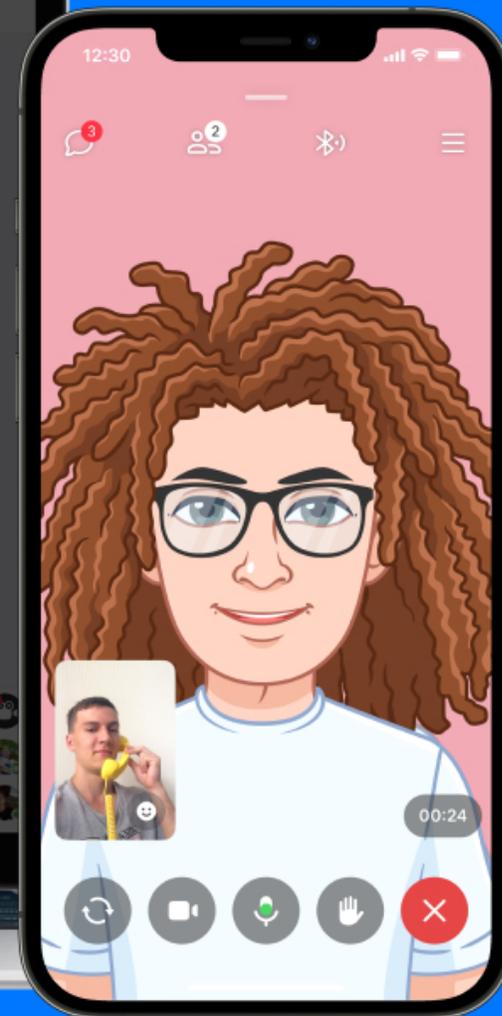
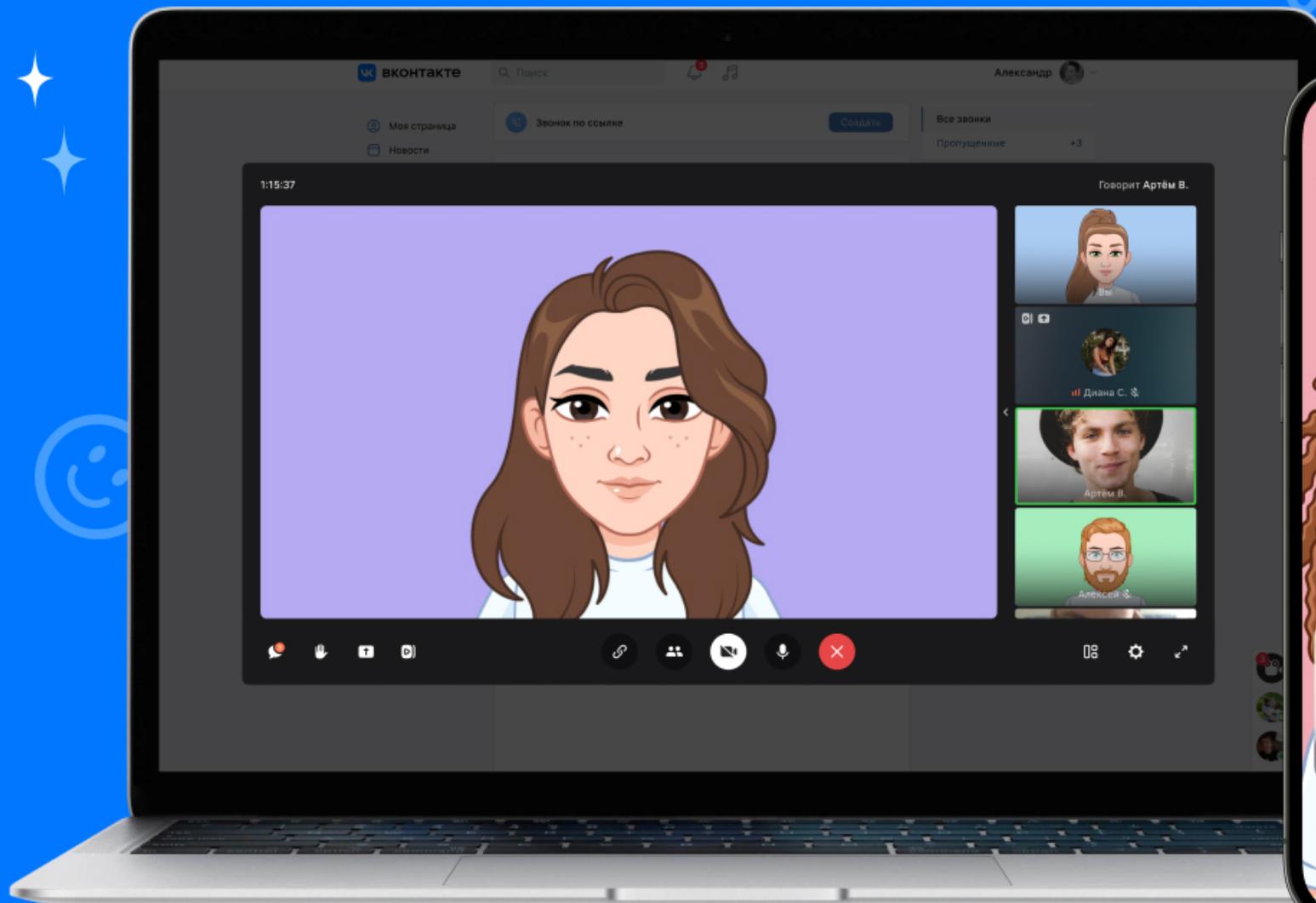
Мой vmoji 



княж Очки Пирсинг **Головной убор** Тело Одежда Экстр

Форма







08:39



Ссылка



★ Вы



Вы



Иван



Больше



Просмотр видео



Показ экрана



Камера



Микрофон



Выйти



Реакции



Участники



Мессенджер



Втожі — ваш представитель в виртуальном мире

- Втожі можно использовать в качестве стикеров в переписке
- Втожі может участвовать в звонке
- Своего персонажа можно редактировать

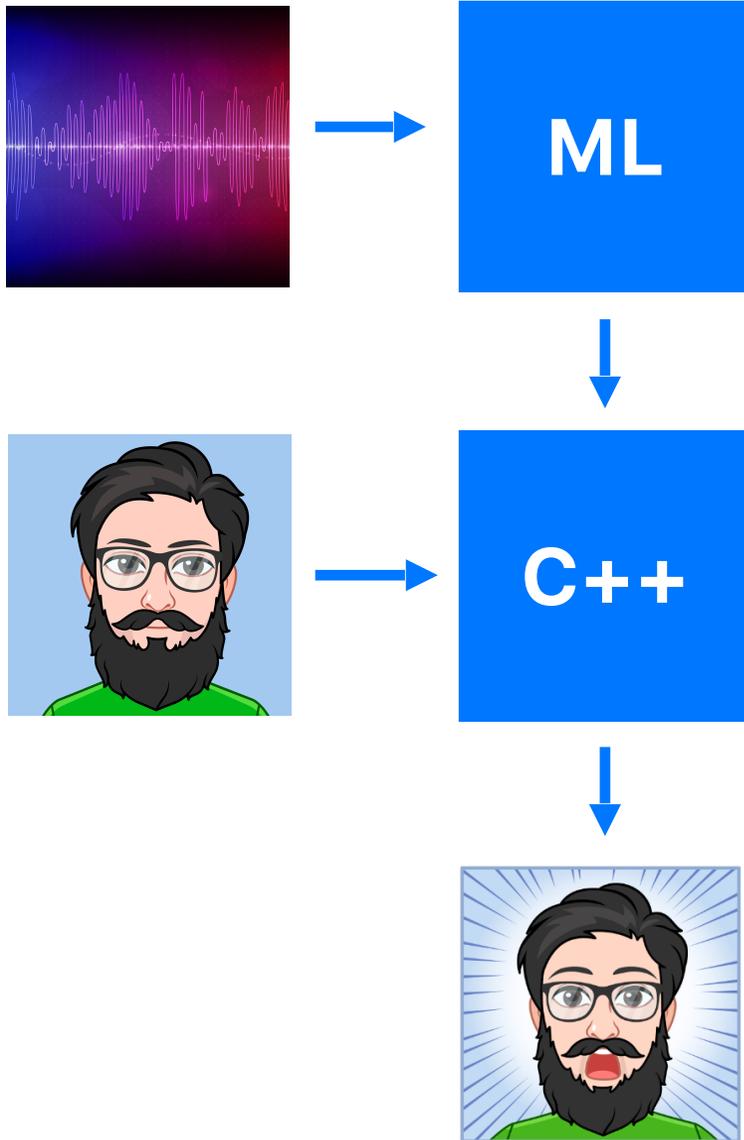
Как реализована анимация





Голос пользователя

- Для анимации рта используется звук с микрофона
- Изображение с камеры не используется — её можно не включать
- Покачивание головой и моргание глаз не зависят от действий пользователя



Общая схема

- Звук с микрофона подается в нейронную сеть
- На выходе сеть дает информацию о движении губ и рта пользователя
- Анимация рендерится на основе статичной картинки и данных от нейросети



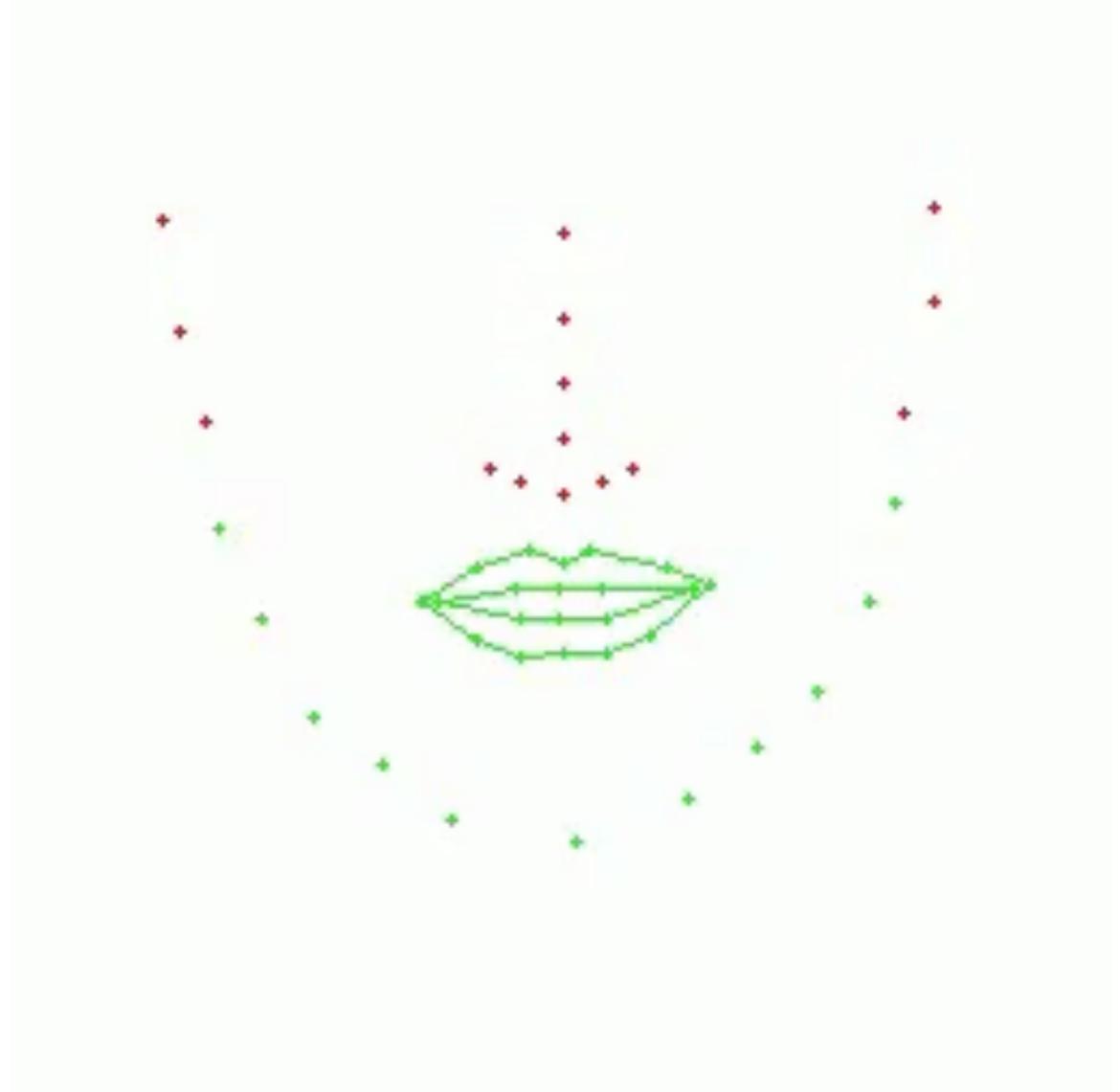
ССЫЛКА НА ПРОЕКТ
WAV2LIP

Машинное обучение

- Вдохновлялись проектом Wav2Lip
- Анимлируем SVG картинку, а не видео
- Нам необходима работа в real time

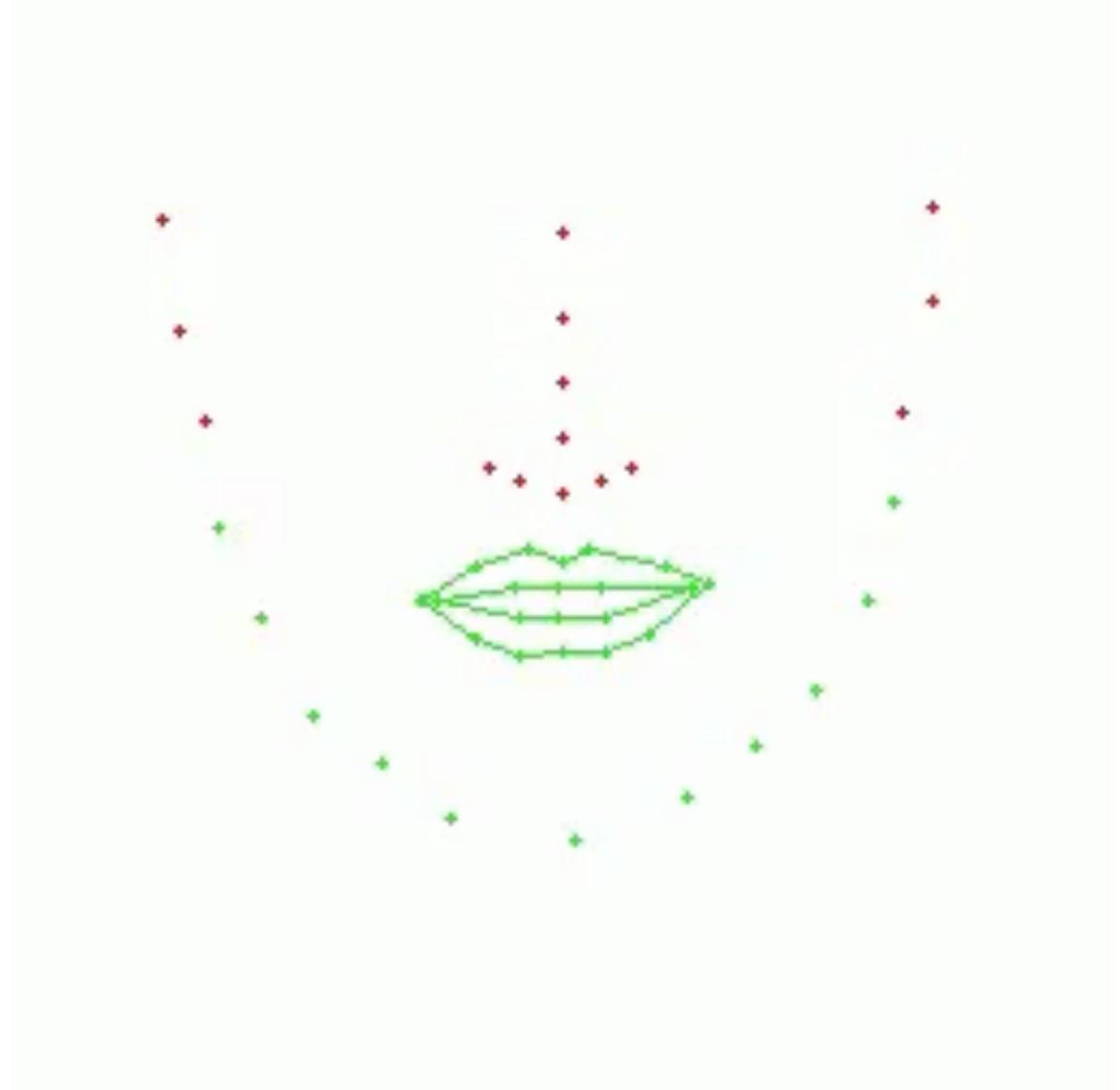
Вход нейросети

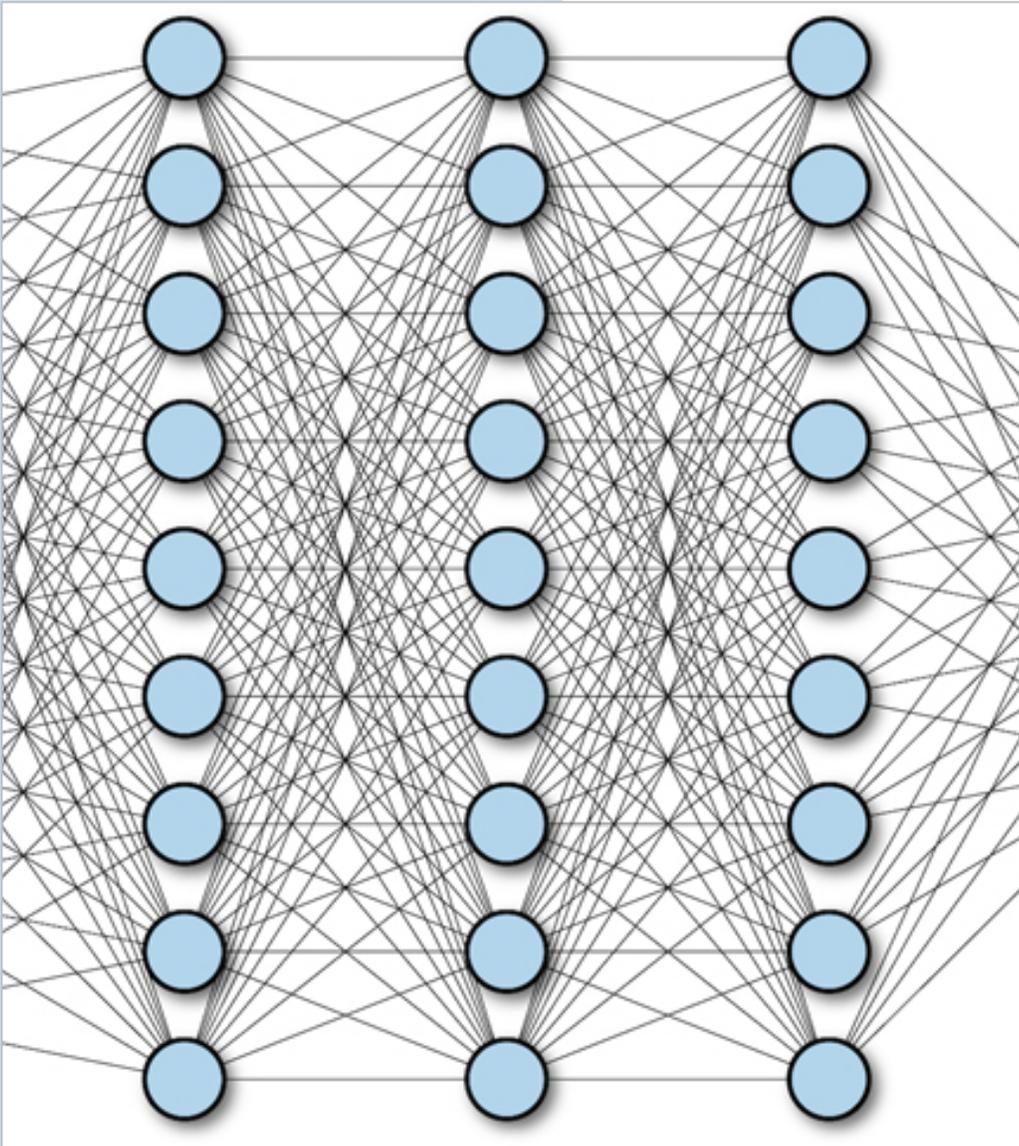
- Верх лица (красные точки)
- Звук с окном в 200 мс
- Вектор спикера
- Рекуррентный вектор для ОС



Вход нейросети

- Низ лица (зеленые точки)
- 31 точка на кадр
- 62 float на кадр
- 25 кадров в секунду





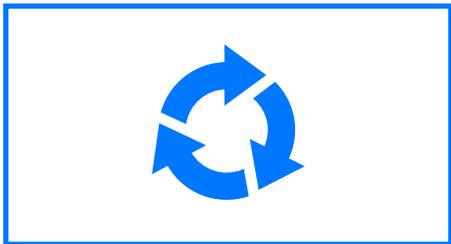
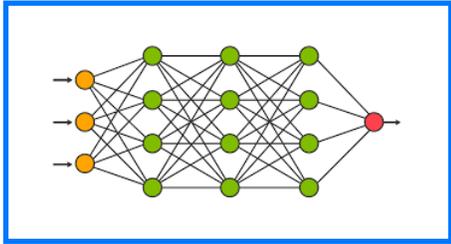
Требования к быстродействию нейросети

- Нейросеть должна работать в клиентском приложении
- В том числе на мобильных платформах
- Работа в режиме реального времени

Характеристики нашей нейронной сети

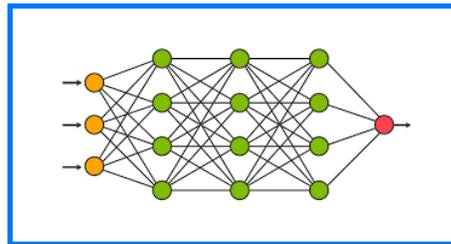
Формат нейросети	TensorFlow Lite (.tflite)
Размер файла нейросети	~ 4 Mb
Время вызова на Snapdragon 662	2 мс
Время вызова на i7 10th gen	0.9 мс

Клиент



Сервер

Клиент



Сервер

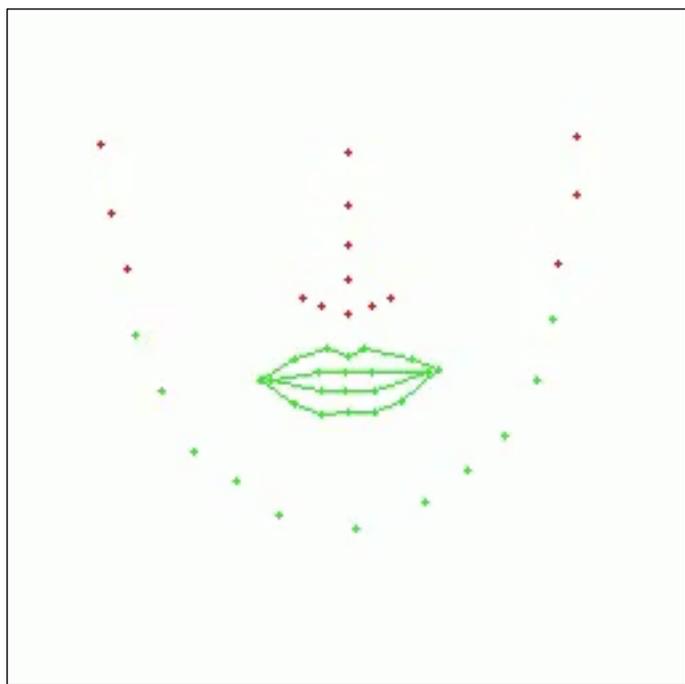
Где запускать нейросеть?

- Со звуком лучше работать ближе к его источнику
- Сетевая задержка
- Искажения из-за потерь в сети



Доклад про звук в
VK Звонках на VideoTech 2022

Отрисовка анимации по полученным точкам



+



=





Рендер на клиенте

- Рендер из SVG
- Изображение высокого качества
- Нет артефактов сжатия
- Требуется разработка клиентского приложения



Рендер на сервере

- Не требуется разработка на стороне клиента
- Присутствуют артефакты кодека
- Степень сжатия и разрешение картинки зависит от пропускной способности сети



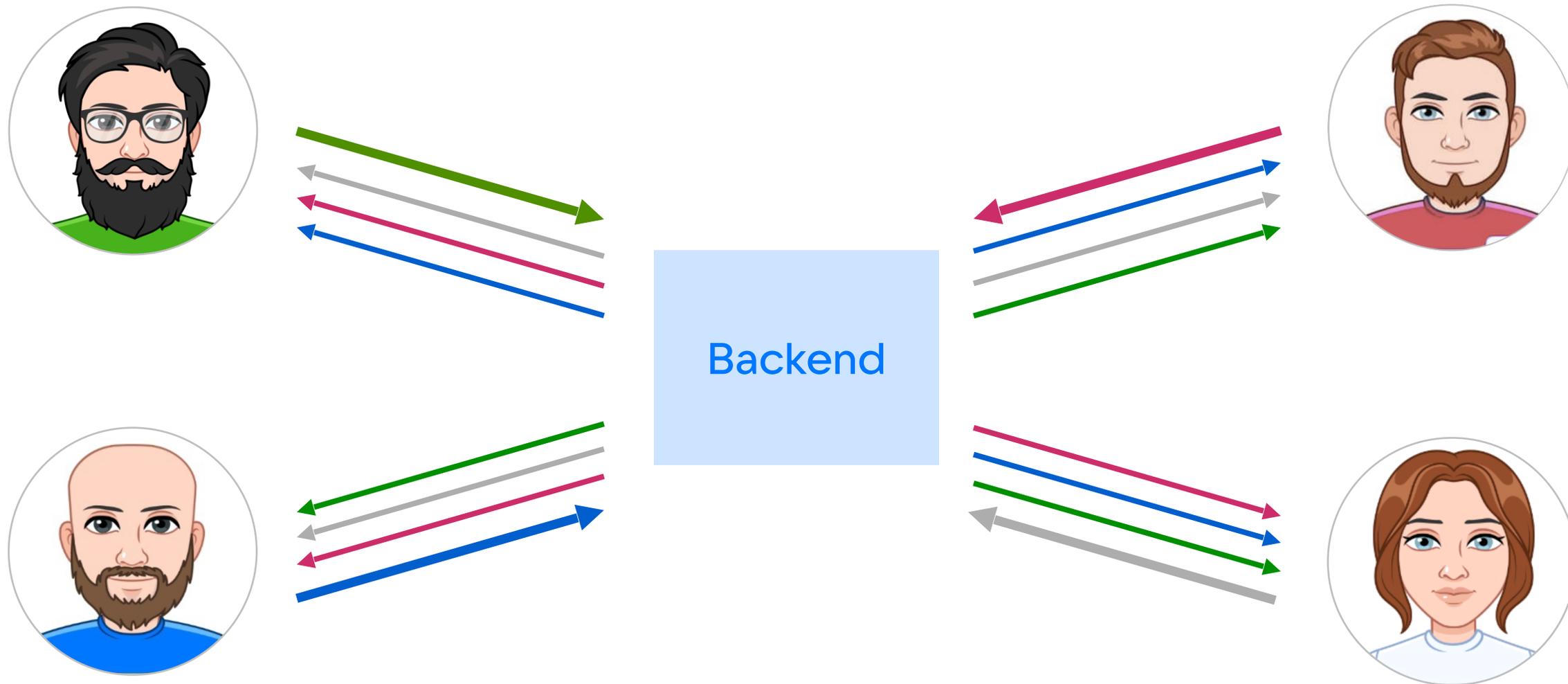
Анимация

- Для перевода звука в анимацию используется нейронная сеть
- На выходе нейросети — координаты точек для анимации
- Анимация рендерится из картинки SVG
- Возможен рендер как на клиенте, так и на сервере
- Среднее время рендера кадра на сервере — 13 мс

Передача данных между пользователями

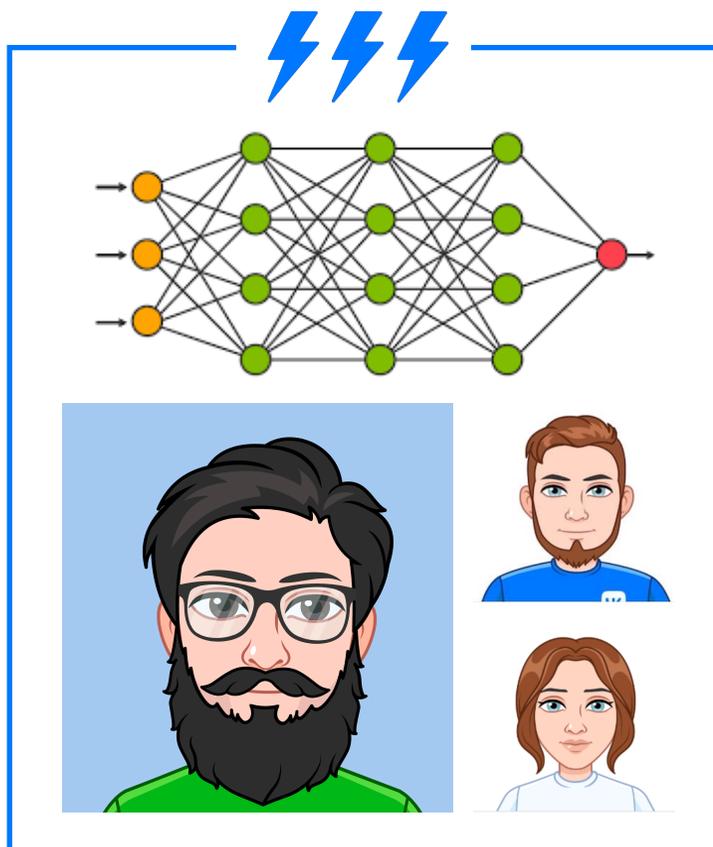


Передача точек между участниками

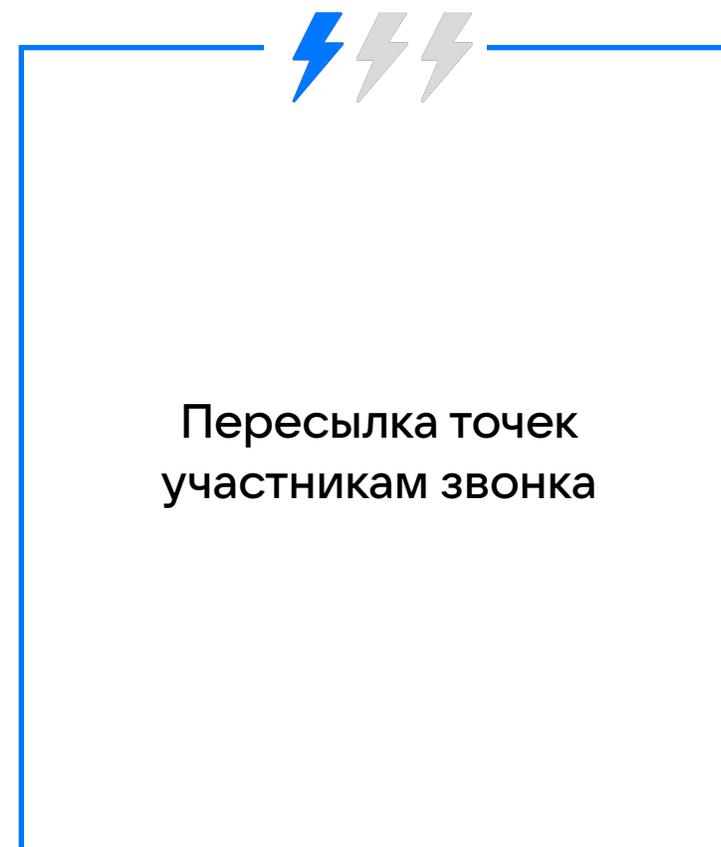


Функции клиента и сервера более подробно

Клиентское приложение



Сервер



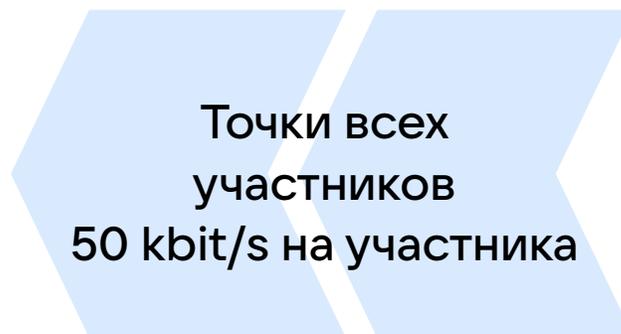
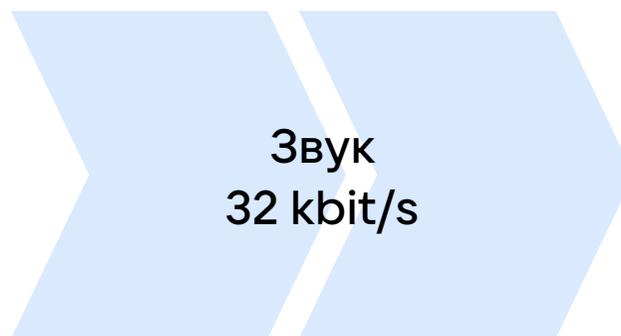
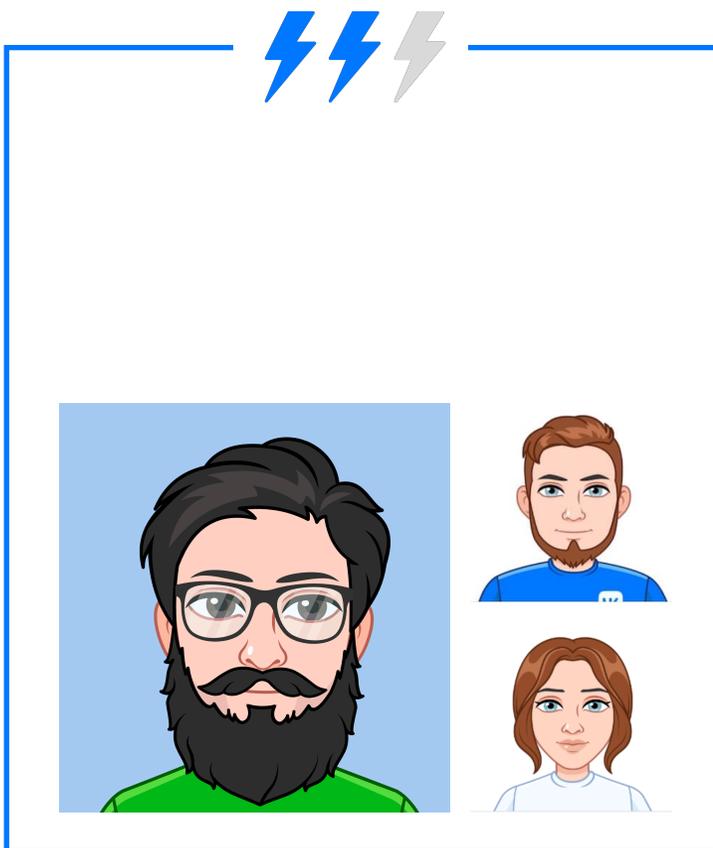
Свои точки
50 kbit/s

Точки других
участников
50 kbit/s на участника

Пересылка точек
участникам звонка

Нейронная сеть на сервере

Клиентское приложение



Сервер



Полный фолбэк на сервер

Клиентское приложение

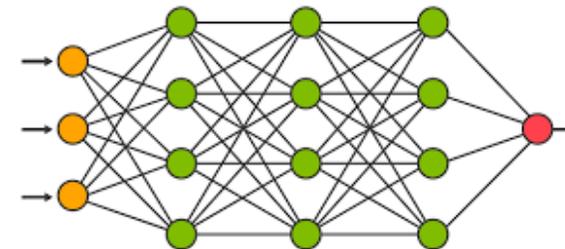


Legacy-клиент

Звук
32 kbit/s

Видео с анимацией
120 – 1000 kbit/s
на участника

Сервер





Как-то так оно всё и работает

- Все участники передают свои точки на сервер
- Сервер раздает точки всем участникам
- Архитектура позволяет сохранить обратную совместимость со старыми версиями клиентов

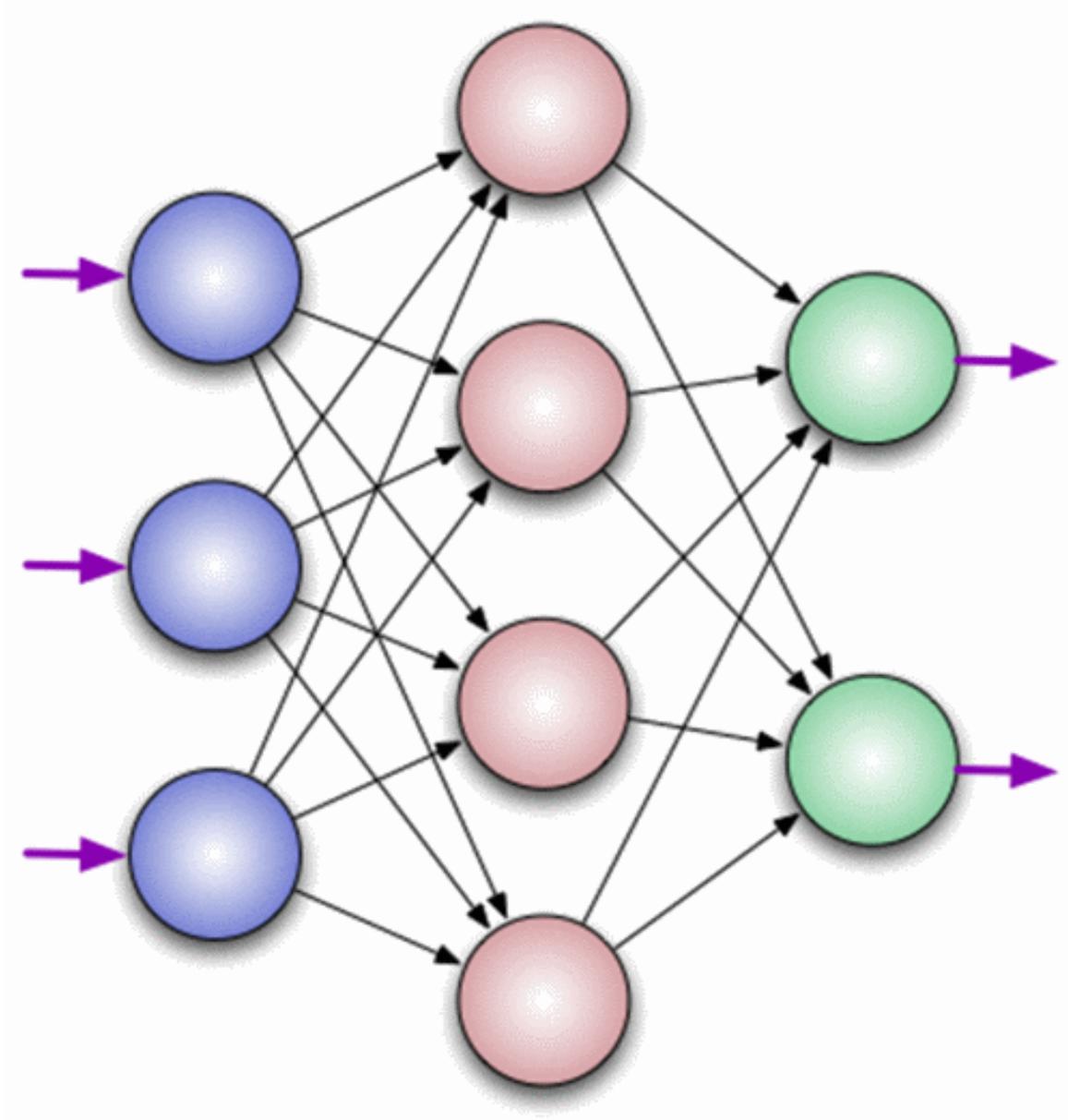
Обо что мы «СПОТКНУЛИСЬ»





Перегенерація vtoji

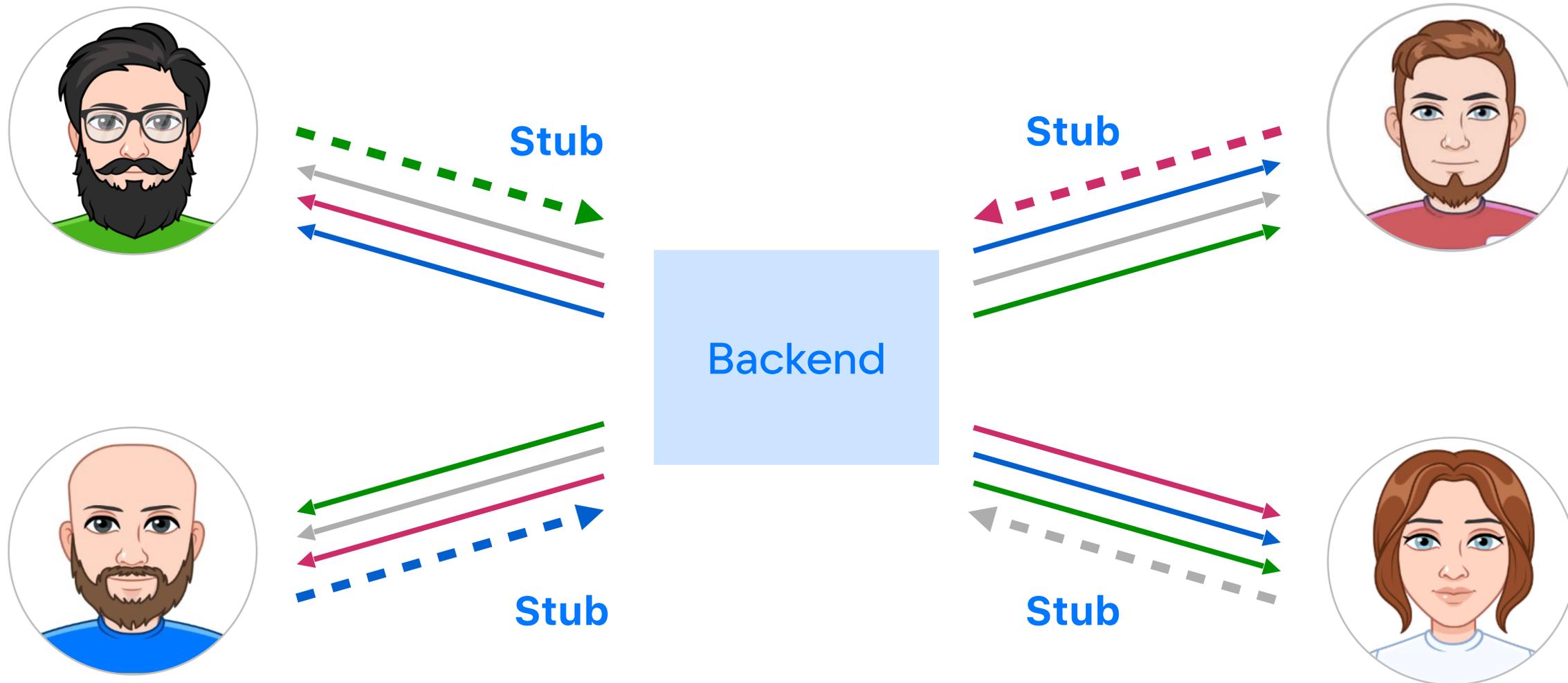
- Продуктове вимога оновити vtoji в дзвінку, якщо користувач його відредагував
- Генерація картинок знаходиться на одному сервері, а дзвінки на іншому
- Потрібно додати колбек про зміну аватара



Нейронная сеть на сервере

- Не учли, что если у клиента сильно ограничены ресурсы CPU, то запускать нейронную сеть нужно на бэкенде
- Пришлось делать запуск нейронной сети на бэкенде в сжатые сроки

Ранняя интеграция





Про что стоит не забыть в начале разработки фичи

- Загруженность всех вовлечённых команд
- Ранняя интеграция
- Обратная совместимость
- Полнота требований

Спасибо
за внимание!

Q&A

