



Asterisk - распознавание и синтез речи.

Речевые технологии и их применение.



Задача вебинара

Рассмотреть реальные кейсы
применения речевых технологий



Предложить и рассмотреть
технические решения для
реализации интеграций

Кейсы

Синтез коротких фраз

Задача: Во время разговора
сообщить важную
персонализированную информацию.

Сложность: минимальная

Инструменты: FastAGI + TTS API

Качество речи по звуку

Задача: найти вызовы с низким качеством связи, не имея данных по RTP и RTCP

Решение: выделение нескольких классов искажений, оценка их влияния на итоговую слышимость звука



Разделение речи

Задача 1: по моно сигналу разделить
речь нескольких человек

Задача 2: идентифицировать их

Решение: выделение
индивидуальных характеристик в
голосе, сохранение известных

Инструменты: chan_rtp + GRPC



Речевая аналитика

Задача:

- контроль голосового модуля
- поиск конфликтов
- кластеризация разговоров
- распознавание языка

Сложность: можно сделать и самому

Инструменты: MixMonitor + STT API

Shazam-like CPA

Задача: обнаружить в RTP стандартные автоответчики, освободить каналы, сэкономить на трафике и операторах

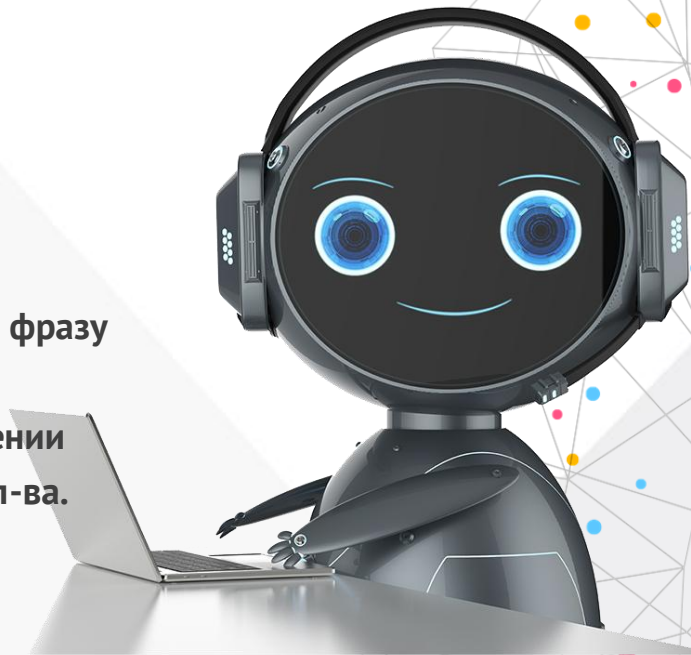
Инструменты: kamailio + rtpengine

Сложносоставные IVR как способ разгрузить call center

- синтез дорого
- распознавание дорого
- чат боты долго и муторно
- повторяющиеся фразы лучше записать
- интеграция с CRM, billing

Зная номер А и причину звонка клиента, полученную с CRM или биллинга, можно проиграть ему записанную фразу тем самым “решив” проблему клиента.

Пример, вы обслуживающая организация, при отключении электроэнергии диктовать что на вашем объекте нет эл-ва.



Голосовые боты

из чего состоит:

- Распознавание речи
- Синтез речи
- Чат-боты

Для чего:

- решение простых рутинных задач
- Экономия на чел/час

Нужно ли вам:

- если в вашем бизнес процессе множество повторяющихся задач.



Готовое ПО

CMU Sphinx

- классическое решение
- HMM
- подходит для изучения и академических целей
- готовые наборы языковых и акустических моделей от NSU

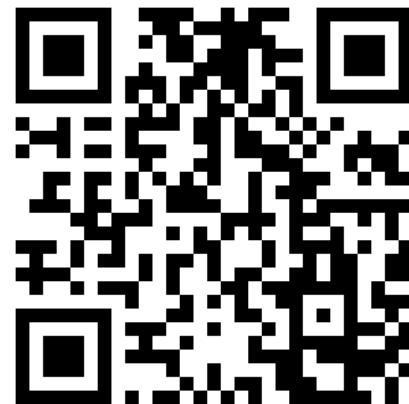
Kaldi

Сборка в docker от alphacephei

- `websocket`
- `python/php/js/c#`
- `asterisk`

Плюсы:

- OpenSource
- Многоязычный



ДЕМО

Запуск kaldi



VOXLINK: ASTERISK
ПРОФЕССИОНАЛЬНО

Yandex SpeechKit



Yandex
SpeechKit

- Первый общедоступный сервис
- 3 режима работы

Плюсы:

- Ф3-152
- Скорость работы



VOXLINK: ASTERISK
ПРОФЕССИОНАЛЬНО

Tinkoff VoiceKit

- Собственная разработка

Плюсы:

- Отличное качество для телефонной речи
- 2 режима работы
- Ценообразование

Минусы: скорость, документация, кредитная оплата



Примеры других систем распознавания

- Готовые:
 - ЦРТ
 - wit.ai
 - google
 - IBM
 - AMAZON
- сделай сам
 - tensorflow
 - deepspeech





ДЕМО

<https://openapi.iqtek.ru/stt/>



VOXLINK: ASTERISK
ПРОФЕССИОНАЛЬНО



ДЕМО

Запись: алло да ну подъезжайте к двум углу до свидания

Yandex: орел да полежать

Tinkoff: Алло! Да. Ну, подъезжайте. Угу. До свидания.

Kaldi: у а ряба на подъезжать

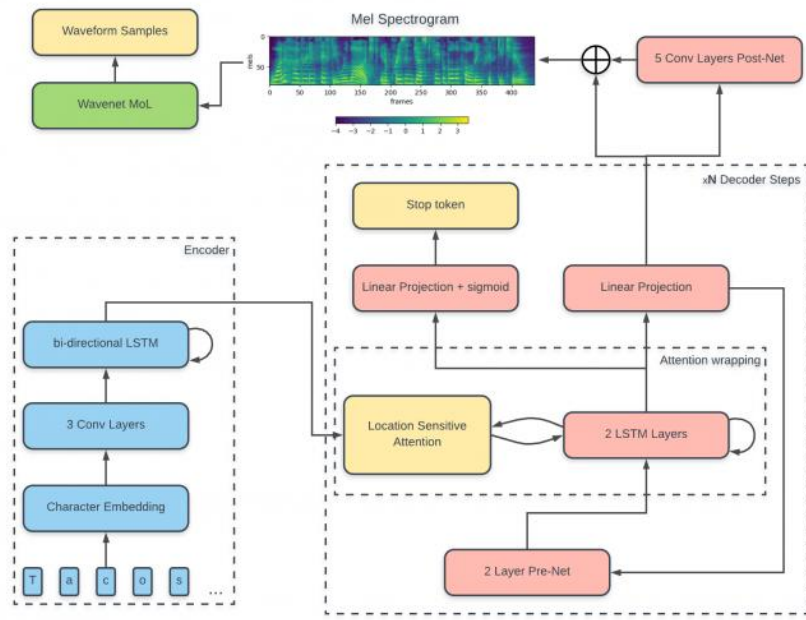
Готовое ПО синтез речи



Yandex

- Премиум голоса
- Тарификация по символам
- Эмоции/ударения

Wavenet



- tinkoff
- yandex
- прочие

Прочие задачи речевых технологий

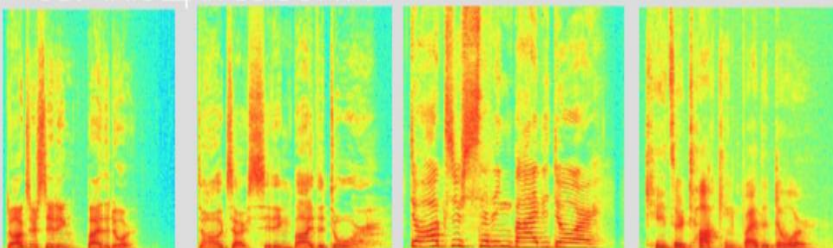


Решаемые задачи

Всё что можно услышать ухом и ощутить -
потенциально решаемые задачи

Эмоциональный окрас речи

ОБУЧАЮЩАЯ ВЫБОРКА

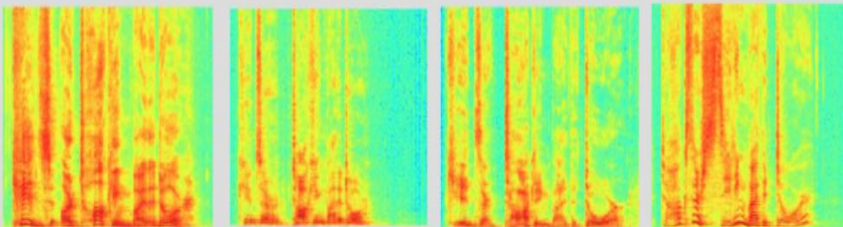


neutral

calm

happy

sad



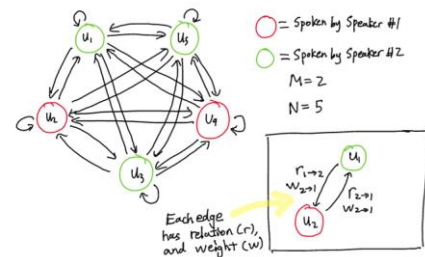
angry

fearful

disgust

surprised

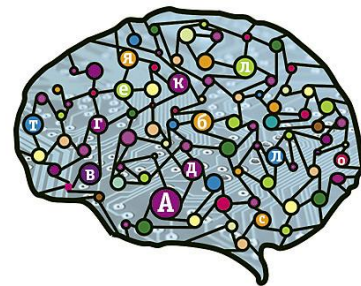
- Нейтральная
- спокойная
- счастливая
- грустная
- злая
- напуганная
- отвращение
- удивление



Семантический анализ речи

Семантический анализ — этап в последовательности действий алгоритма автоматического понимания текстов, заключающийся в выделении семантических отношений, формировании семантического представления текстов. Один из возможных вариантов представления семантического представления — структура, состоящая из «текстовых фактов».

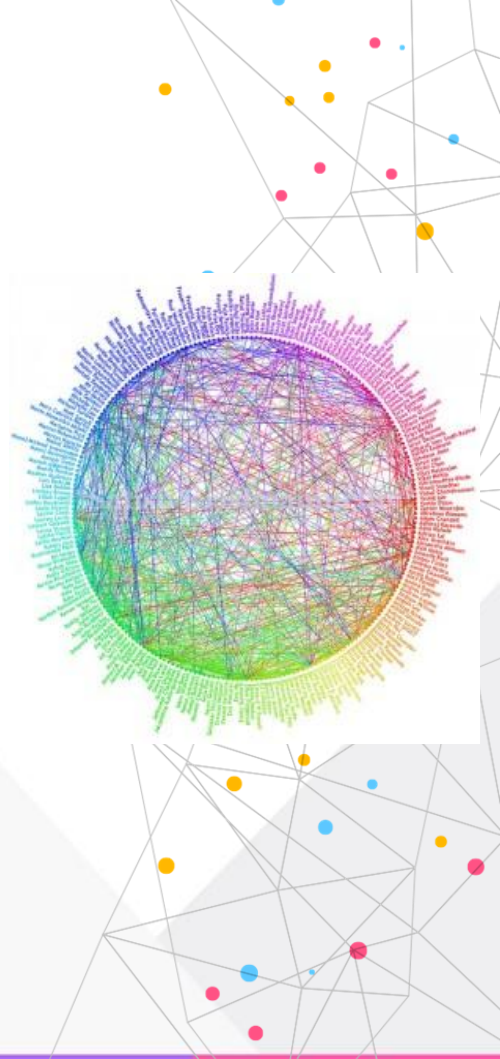
Семантический анализ в рамках одного предложения называется локальным семантическим анализом.



Семантический анализ речи

Примеры:

1. Иголки Ёжика совсем не напоминают швейные иголки.
2. Согласие:
 - a. Да хорошо
 - b. ой, ну давайте
 - c. ну, наверное да
 - d. без базара, братан.
 - e. ну чего думать то, конечно
 - f. ну что ж выбора нет, везите
3. ДОРОГО



Биометрия голоса

Надежность:

FRR (False Rejection Rate) – вероятность ложного отказа в доступе (ошибка первого рода)

FAR (False Acceptance Rate) – вероятность ложного допуска, когда система ошибочно опознает чужого как своего (ошибка второго рода).

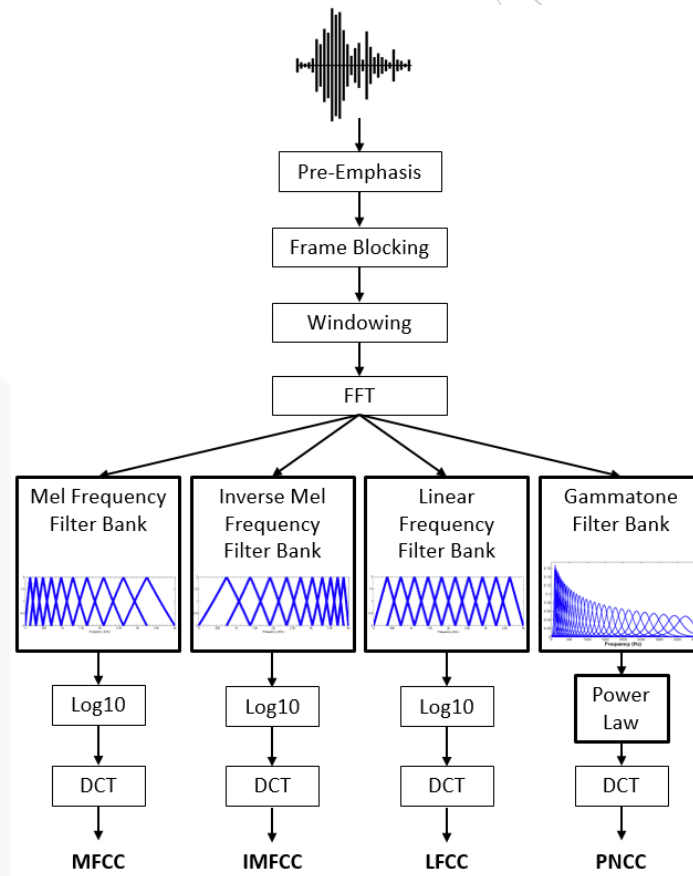
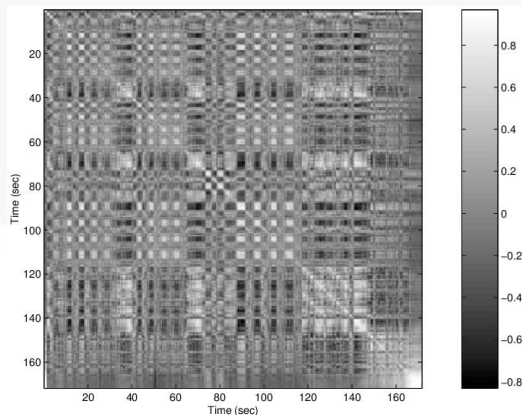
- Не может дать 100% результат!

Биометрический признак	Тест	Условия тестирования	FRR %	FAR %
Отпечатки пальцев	FVC 2006	Неоднородная популяция (включает работников ручного труда и пожилых людей)	2,2	2,2
Лицо	MBE 2010	Полицейская база фотографий	4,0	0,1
		База фотографий с документов	0,3	0,1
Голос	NIST 2010	Текстонезависимое распознавание	3..4	1,0
Радужная оболочка глаз	ICE 2006	Контролируемое освещение, широкий диапазон качества изображений	1,1...1,4	0,1

Биометрия голоса

MFCC:

- Мел-кепстральные коэффициенты
- Кепстральный анализ
- https://github.com/agicblack/voice_bio



Биометрия голоса

Распознавание по голосу – одна из форм биометрической аутентификации, позволяющая идентифицировать личность человека по совокупности уникальных характеристик голоса(c)wiki

Задачи биометрии:

1. Определения пола
2. Определения возраста
3. Аутенфикация



Решаемые задачи (примеры)

- Извлечение метаинформации
- Изменение стиля речи
- Анализ видео (deepfake)

Примеры и ссылки по завершению

Интеграция



Методы

По актуальности:

- Устаревающие
- Нативные
- Сторонние
- Экзотика

По применимости:

- В реальном времени:
 - Перехват голоса
 - Встраивание голоса
- Пост-обработка

Методы: UniMRCP и EAGI

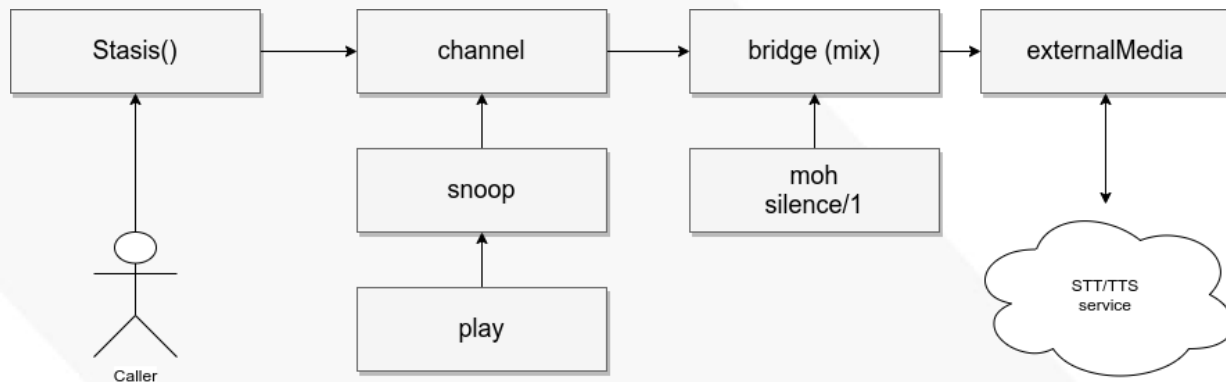


Методы: MixMonitor (stereo) + ChanSpy

- **MixMonitor (постобработка)**
 - **Запись в стерео**
 - **Выбор bitrate**
 - **Последующая обработка**
- **ChanSpy (перехват в реальном времени)**
 - **передача RTP**
 - **Dial(UnicastRTP/127.0.0.1:10000)**



Методы: ARI





ДЕМО

бот по номеру телефона +7(499)380-90-47



VOXLINK: ASTERISK
ПРОФЕССИОНАЛЬНО

Методы:

- Tinkoff
 - GRPC
- rtpengine-recording
 - TLS сокет
- voipmonitor
 - Собственное API
- SIPREC
 - Поддерживается OpenSIPS

Ваши вопросы?





ДЕМО

Принимаем вопросы в онлайн
телефон: +7(499)380-90-48



VOXLINK: ASTERISK
ПРОФЕССИОНАЛЬНО



С чего начать?

- Курсы + Контесты
- Статьи
- Sphinx + Kaldi
- Youtube (nsu)
- Общение в телеграм
- Wavenet + tensorflow

или интегратор

Лучше иметь реальную задачу и данные для исследования и верификации

Как воспринимают кожаные мешки?

- Не скрываем, что мы - робот
- Дополнять операторов
- Во время ожидания - пробовать ответить на вопросы
- Если человек будет “тупить” и переспрашивать - он тоже не понравится

Когда IVR лучше?

- Когда окупается
- Для исходящих
 - DTMF плохо воспринимается
 - Лучше общение голосом



Существуют сервисы ИИ?

- Универсальных нет
- То что максимально близко - дорого
 - IBM
- Алиса
 - Не ИИ

Применение в такси

- Датасет пьяных пассажиров
 - \$\$\$
- Свои приложения и передача речи не через ТФОП - лучшее распознавание

Ваши вопросы?

Вопросы
из онлайн





СПАСИБО ЗА ВНИМАНИЕ!

Замятин Михаил

tg:@agick

Грушко Сергей

+7 (926) 350 3163

sergey@voxlink.ru

Гончаровский Игорь

tg:@igorrg

igorg@iqtek.ru