ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

ZIAX TTS: Синтез речи

ИНСТРУКЦИЯ ПО ЭКСПЛУАТАЦИИ

(Руководство разработчика)

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

Содержание

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

Аннотация к описанию API GRPC сервера системы «ZIAX TTS»

Задачи настоящей документации:

Настоящая документация необходима для быстрой адаптации существующих и новых продуктов в компании с системой синтеза речи «ZAIX TTS». Данная документация содержит информацию, необходимую для эксплуатации экземпляра ПО «ZAIX TTS: Синтез речи».

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

1. API сервиса

1.1 Версия протокола

Текущая версия API: v1.

1.2 Файл «.proto»

```
syntax = "proto3";

package ziax.tts.v1;

message RawAudio {
  enum AudioEncoding {
    AUDIO_ENCODING_UNSPECIFIED = 0;

    // Audio bit depth 16-bit signed little-endian (Linear PCM)
    LINEAR16_PCM = 1;
  }

  // Encoding type
  AudioEncoding audio_encoding = 1;

  // Sampling frequency of the signal
  int64 sample_rate_hertz = 2;
}

message ContainerAudio {
  enum ContainerAudioType {
    CONTAINER_AUDIO_TYPE_UNSPECIFIED = 0;

    // Audio bit depth 16-bit signed little-endian (Linear PCM)
    WAV = 1;

    // Data is encoded using the OPUS audio codec and compressed using the OGG container format
    OGG_OPUS = 2;

    // Data is encoded using MPEG-1/2 Layer III and compressed using the MP3 container format
    MP3 = 3;
  }

  ContainerAudioType container_audio_type = 1;
}

message AudioFormatOptions {
  oneof AudioFormat {
    // The audio format specified in request parameters
    RawAudio raw_audio = 1;

    // The audio format specified inside the container metadata
    ContainerAudio container_audio = 2;
```

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

```
    }
}

message AudioChunk {
    // Sequence of bytes of the synthesized audio in format specified in output_audio_spec
    bytes data = 1;
}

message Hints {
    // The hint for TTS engine to specify synthesized audio characteristics
    oneof Hint {
        // Name of speaker to use
        string voice = 1;

        // Hint to change speed
        double speed = 2;

        // Hint to specify pronunciation character for the speaker
        string role = 3;
    }
}


message UtteranceSynthesisResponse {
    // Part of synthesized audio
    AudioChunk audio_chunk = 1;
}

message UtteranceSynthesisRequest {
    // The name of the model
    string model = 1;

    // Text to synthesis, one of text synthesis markups
    oneof Utterance {
        // Raw text (e.g. "Hello world!")
        string text = 2;
    }

    // Optional hints for synthesis
    repeated Hints hints = 3;

    // Optional. Default: 22050 Hz, linear 16-bit signed little-endian PCM, with WAV header
    AudioFormatOptions output_audio_spec = 4;
}


// A set of methods for voice synthesis
service TtsService {
    // Synthesizing text into speech
    rpc UtteranceSynthesis (UtteranceSynthesisRequest) returns (stream UtteranceSynthesisResponse) {
    }
}
```

Актуальную версию proto-файла можно запросить у менеджера проектов.

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

1.3 Описание сервиса – протокола

Иерархия типов:

- UtteranceSynthesisRequest
  - Hints
  - AudioFormatOptions
    - RawAudio
    - ContainerAudio
- UtteranceSynthesisResponse
  - AudioChunk

1.3.1 UtteranceSynthesisRequest

Запрос на синтез речи:

Поле	Тип	Описание
model	string	Название используемой модели Допустимое значение: general
text	Utterance	Синтезируемый текст
hints	Hints []	Список характеристик синтезируемого аудио (множество) Опциональное поле
output_audio_spec	AudioFormatOptions	Параметры спецификации синтезируемого аудио Опциональное поле

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

1.3.2 Hints

Характеристики синтезируемой речи:

Поле	Тип	Описание
voice	string	Имя спикера (название голоса)
speed	double	Скорость воспроизведения речи Значение по умолчанию: 1.0
role	string	Амплуа голоса Значение по умолчанию: neutral

1.3.3 AudioFormatOptions

Параметры синтезируемого аудио:

Поле	Тип	Описание
raw_audio	RawAudio	Параметры получения аудио контента без сжатия и заголовка
container_audio	ContainerAudio	Параметры получения аудио контента внутри контейнера (файла)

1.3.4 RawAudio

Параметры аудиоданных без сжатия и заголовка:

Название	Тип	Описание
audio_encoding	AudioEncoding	Тип кодирования
sample_rate_hertz	int64	Частота дискретизации синтезируемого аудио

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

1.3.5 AudioEncoding (перечисление)

Формат аудиокодирования:

Название	Значение	Описание
AUDIO_ENCODING_UNSPECIFIED	0	Формат не установлен
LINEAR16_PCM	1	Без заголовка (Linear PCM/16-bit signed)

1.3.6 ContainerAudio

Параметры контейнера с аудиоданными (полностью готовый файл):

Название	Тип	Описание
container_audio_type	ContainerAudioType	Формат аудиоконтейнера

1.3.7 ContainerAudioType (перечисление)

Формат аудиокодирования:

Название	Значение	Описание
CONTAINER_AUDIO_TYPE_UNSPECIFIED	0	Формат не установлен
WAV	1	С заголовком (Linear PCM/16-bit signed little-endian)
OGG_OPUS	2	Данные кодируются аудиокодеком OPUS и сжимаются в контейнер формата OGG
MP3	3	Данные кодируются аудиокодеком MPEG-1/2 Layer III и сжимаются в контейнер формата MP3

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

1.3.8 UtteranceSynthesisResponse

Ответ с результатами синтеза речи:

Поле	Тип	Описание
audio_chunk	AudioChunk	Фрагмент аудио данных

1.3.9 AudioChunk

Фрагмент синтезируемых аудио данных:

Поле	Тип	Описание
data	byte []	Массив байт с аудио контентом в формате заданным параметром запроса: output_audio_spec

1.3.10 Расположение сервиса

В docker окружении, сервис расположен по адресу: tts:5001, где tts — имя контейнера, а 5001 — TCP порт подключения

1.3.11 Использование сервиса

Сервис предназначен для синтеза речи в режиме реального времени
Для его использования, необходимо создать собственное клиентское приложение, предварительно сгенерировав код интерфейса из соответствующего *.proto файла

Реализация клиента выполняется по следующим правилам:

- Установка соединения (Channel) с gRPC сервером
- Создание потока (Stub), который будет обслуживать очередь выполнения для запросов и ответов (CompletionQueue)
- Отправка запроса с параметрами конфигурации и текстом для синтеза речи
- Получение результатов синтезируемого аудио
- Закрытие потока
- Закрытие соединения

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

ВАЖНО

- Параметр output_audio_spec может содержать параметры конфигурации в формате RawAudio или ContainerAudio, но не в обоих форматах одновременно

gRPC-соединение (Channel), поддерживает мультиплексирование потоков (Stub). Клиентское приложение должно быть спроектировано так, чтобы в рамках одного соединения параллельно обрабатывать до 100 потоков одновременно. При достижении "лимита" на количество открытых потоков (превышение лимита предполагает потери в производительности), рекомендуется создать новое соединение с уникальными ChannelCredentials, т.е. с отличными параметрами, от уже созданного соединения, иначе параметры нового соединения будут скопированы из текущего и новое соединение создано не будет

Поток (Stub), использует двунаправленную потоковую передачу RPC ([bidirectional / bidi] streaming RPC), позволяя тем самым сделать потоки клиент/сервера независимыми, но накладывает дополнительные ограничения на передачу данных

После создания потока, первым и единственным запросом, клиент передает формат результирующего аудио (output_audio_spec), устанавливает дополнительные характеристики синтезируемой речи, в виде дополнительных подсказок сервису (hints), такие как голос (voice) и скорость произношения (speed) и т.п., определяет и задаёт текст (text), который необходимо синтезировать

Ответ с результатами синтезируемой речи (Response), содержит один или несколько фрагментов данных (chunks) упакованных в соответствии с требуемым форматом (AudioFormatOptions), определяемым в запросе на синтез речи (Request), полем (output_audio_spec). Если в качестве формата аудио используется тип (RawAudio), то (chunks) будут предоставлять данные без заголовка (Linear PCM/16-bit signed). Если же в качестве формата аудио использовался тип (ContainerAudio), то (chunks) предоставляют данные в виде готового файла

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

2. Примеры

Примеры по созданию клиентов и генерации кода интерфейса, можно найти по адресу https://gitlab.satel.org/ziax/protofiles/-/tree/master/tts/examples или запросить у менеджера проектов.

2.1 Python

2.1.1 Генерация кода интерфейса

Для генерации кода интерфейса, необходимо склонировать репозиторий googleapis.git например в директорию «third_party» или склонировать проект ziax/protofiles и выполнить инициализацию модулей git

```
$ git submodule update --init --recursive
```

Установить пакет «grpcio-tools»

```
$ pip install grpcio-tools
```

Затем сгенерировать код интерфейса

```
$ cd tts/examples
$ python \
    -m grpc_tools.protoc \
    -I../v1/ \
    --python_out=. \
    --grpc_python_out=. \
    ../v1/tts_service.proto
```

Результат генерации кода интерфейса

```
$ ls -1 *pb2*
tts_service_pb2_grpc.py
tts_service_pb2.py
```

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

2.1.2 Пример клиента

```
import argparse
import grpc
import wave

from tts_service_pb2 import AudioFormatOptions, Hints, RawAudio, UtteranceSynthesisRequest
from tts_service_pb2_grpc import SynthesizerStub


def __create_args_parser():
    ap = argparse.ArgumentParser()
    ap.add_argument('-t', '--text', required=True, help='Text to be synthesized'')
    ap.add_argument('-o', '--out-file', required=True, help='Output WAV file'')
    ap.add_argument('-e', '--endpoint', default='127.0.0.1:5001', help='gRPC server endpoint')
    ap.add_argument('-v', '--voice', default='elena', help='Speaker voice identifier')
    ap.add_argument('-s', '--speed', type=float, default=1.0, help='Pronunciation speed')
    ap.add_argument('-r', '--sample-rate', type=int, default=22050, help='Sample rate in Hertz')

    return ap


def __main():
    args = __create_args_parser().parse_args()

    channel = grpc.insecure_channel(args.endpoint)
    stub = SynthesizerStub(channel)
    hints = [
        Hints(voice=args.voice),
        Hints(speed=args.speed)
    ]

    audio_format = AudioFormatOptions(
        raw_audio=RawAudio(
            audio_encoding=RawAudio.LINEAR16_PCM,
            sample_rate_hertz=args.sample_rate
        )
    )

    with wave.open(args.out_file, 'wb') as writer:
        writer.setnchannels(1)
        writer.setsampwidth(2)
        writer.setframerate(args.sample_rate)

        it = stub.UtteranceSynthesis(UtteranceSynthesisRequest(
            text=args.text,
            hints=hints,
            output_audio_spec=audio_format
        ))
        for response in it:
            writer.writeframes(response.audio_chunk.data)


if __name__ == '__main__':
    __main()
```

3. Подключение

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

Следующие разделы текущего параграфа, описывают процесс подключения к серверу распознавания.

3.1 Сетевое подключение

Все данные отправляемые и получаемые от сервера распознавания «Ziax» передаются в обезличенном виде по открытым каналам связи.

3.2 Порт подключения

Для передачи данных («gRPC streaming») используется порт подключения «tcp:5001».

4. Ссылки

- ziax/protofiles
- googleapis.git

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

5. Перечень терминов и сокращений

Термины, сокращения и определения, используемые в настоящем документе, приведены в таблице 2.

Таблица 2 – Термины, сокращения и определения

Термин/ Сокращение	Определение/Пояснение
API	Программный интерфейс приложения, интерфейс прикладного программирования (англ. application programming interface) – описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой. Используется программистами при написании всевозможных приложений
Docker	ПО с открытым исходным кодом для автоматизации развёртывания и управления Программы с использованием технологии контейнеризации. В состав «Docker» включен пакетный менеджер «Docker Compose», обеспечивающий запуск многоконтейнерных приложений
gRPC	Система удалённого вызова процедур (RPC – Remote Procedure Calls) с открытым исходным кодом, разработанная компанией «Google». В качестве транспорта используется протокол HTTP/2, в качестве языка описания интерфейса – буферы протоколов. gRPC предоставляет функции аутентификации, двунаправленной потоковой передачи и управления потоком, блокирующие или неблокирующие привязки, а также отмена и тайм-ауты. Генерирует кроссплатформенные привязки клиента и сервера для многих языков. Наиболее часто используется для подключения служб в микросервисном стиле архитектуры и подключения мобильных устройств и браузерных клиентов к серверным службам
TCP\IP	Сетевая модель передачи данных, представленных в цифровом виде. Модель описывает способ передачи данных от источника информации к получателю. В модели предполагается прохождение информации через четыре уровня, каждый из которых описывается правилом (протоколом передачи). Наборы правил, решающих задачу по передаче данных, составляют стек протоколов передачи данных, на которых базируется Интернет. Название TCP/IP происходит из двух основных протоколов семейства – Transmission Control Protocol (TCP) и Internet Protocol (IP), которые были первыми разработаны и описаны в данном стандарте
TTS	
PCM	
WAV	
OGG	
OPUS	

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

MP3	
MPEG-1/2 Layer III	

ООО «ЗИАКС»
410056, Саратовская область, г. Саратов, ул. Советская, дом 61
+7 (495) 108-70-01
info@ziax.ru; ziax.ru

Лист регистрации изменений								
Изм	*Номера листов (страниц)*			*Всего листов (страниц) в документе*	*№ документа*	*Входящий № сопроводит. документа и дата*	*Подп.*	*Дата*
	Замененных	*Новых*	*Аннулированных*					