

## **ГОЛОСОВОЙ ПОМОЩНИК ДЛЯ ПЛАТФОРМЫ 1С: ПРЕДПРИЯТИЕ**

**Аннотация.** В статье описывается процесс создания голосового помощника для платформы 1С: Предприятие. Представлены методы распознавания и синтеза голоса. Проведен анализ мер сходств. Реализован алгоритм работы голосовых помощников. Обозначены перспективы развития программного приложения.

**Ключевые слова:** голосовой помощник, распознавание речи, синтез речи, метрики сходства, голосовой ассистент.

### **Введение**

Информационное общество предполагает широкое применение различных информационных технологий в повседневной деятельности. На данный момент невозможно представить работу предприятия без информационной системы, которая обеспечивает распространение, хранение и обработку информации.

Самым популярным решением на российском рынке выступают продукты 1С. С накоплением большого количества данных информационные системы становятся неструктурированными, ориентироваться в них становится сложнее. Работа ограничена рутинными действиями, на которые затрачивается основная часть времени.

На сегодняшний день популярным инструментом расширения информационных систем становится внедрение цифровых ассистентов с голосовым управлением. Голосовые помощники, существующие на данный момент, справляются с задачей поиска и обработки информации, а также выполнением определенных действий. Привязать существующего голосового помощника к платформе 1С: Предприятие невозможно. Поэтому для улучшения системы необходимо разработать универсального голосового ассистента,

который будет подключаться к любой информационной базе. Данное внедрение повлияет на работу персонала: появится возможность обращаться к данным из базы голосом, выполнять часть работы на фоне, не переключая внимания с основной деятельности. Кроме того предполагается, что информационная система с голосовым помощником повысит свою конкурентоспособность на рынке с похожими продуктами.

### **Постановка задачи**

В данной работе была поставлена цель – создание голосового помощника для платформы 1С: Предприятие на языке python с использованием библиотек. Для достижения поставленной цели необходимо решить ряд задач:

- разработать алгоритм работы голосовых помощников;
- проанализировать методы распознавания голоса;
- сравнить меры сходства для текста;
- изучить алгоритмы синтеза речи;
- разработать программное приложение.

В результате получим приложение, удовлетворяющее заявленным требованиям.

### **Алгоритм работы системы распознавания**

Голосовые помощники, существующие на сегодняшний день, имеют одинаковый принцип работы, различия состоят лишь в механизмах, которые задействованы на том или ином этапе (рис. 1). Сначала произнесенная человеком речь попадает в модуль распознавания.



*Рис. 1.* Алгоритм распознавания речи

На следующем этапе преобразованная в текстовый запрос речь проходит классификацию по интентам. Под интентом в информационном поиске подразумевают ключевые слова запроса. Выделяют несколько основных интенгов исходя из предполагаемого целевого действия:

- 1) получение информации;
- 2) выполнение определенных навыков (например, поставить напоминание, открыть приложение, заказать такси);
- 3) выполнение действий системы (например, прибавить/убавить громкость, выключиться, активироваться).

После определения интенга, текстовый запрос подвергается семантическому анализу. На данном этапе выделяются отношения между словами, происходит понимание текста в рамках интенга. Результаты разбора превращаются в некоторое структурированное представление, воспринимаемое программой, которое называется фреймом.

Данные из фрейма подставляются в алгоритм действий, получаемый из некоторой базы. Алгоритм попадает в модуль выполнения, который, помимо этого, знает текущий контекст диалога, знает, что происходило до этого момента. На основании выполненных действий генерируется текст, который попадает в модуль синтеза речи. Сформулированный ответ слышит пользователь.

### **Распознавание голоса**

Распознавание речи проходит следующие этапы:

1. Отделение шума от голоса (рис. 2).

Существует несколько методов для выполнения этого этапа:

- выделение признаков, которые являются инвариантными относительно шума;
- использование алгоритмов, логика которых построена на коэффициентах линейного предсказания, мел-кепстральных коэффициентах (коэффициенты, характеризующие сигнал на основе его спектра и амплитуды);
- проведение цифровой обработки (маскирование шума, шумоподавление с использованием нескольких микрофонов). Под маскированием шумов

понимаются алгоритмы, при которых численные значения, которые оказываются сравнимыми с характеристиками шума либо игнорируются системой, либо получают меньшие весовые коэффициенты;

- построение массивов микрофонов. Данная технология позволяет выделить направленный микрофон (наиболее значимый) с переменным лучом направления.

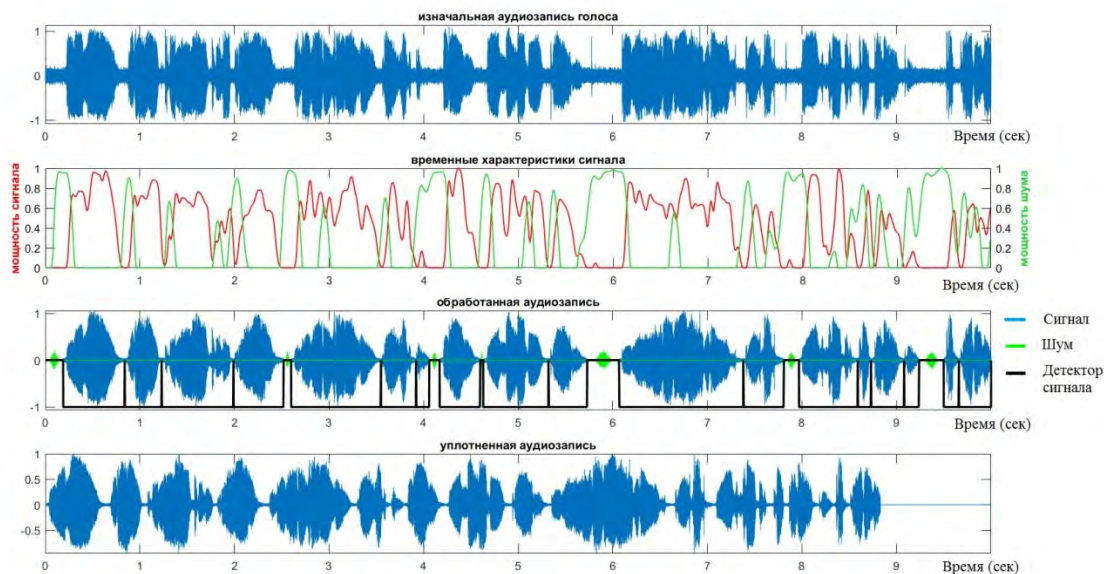


Рис. 2. Этапы отделения шума от голоса

## 2. Преобразование звука в акустические параметры.

Звуковой сигнал разбивают на отрезки одинаковой длины. Затем с помощью дискретных преобразований, например, преобразование Фурье, преобразуют в частотную область.

## 3. Приведение акустической формы сигнала к фонемам.

Разбиение акустических параметров на области сгущения, для выделения элементов фонем. Полученные элементы сравниваются с эталонами. В качестве эталонов в словарях хранятся фонемные элементы (в зависимости от полноты словаря, могут храниться как фонемы слов, так и отдельных слогов).

## 4. Определение последовательности фонем и преобразование к тексту.

Существует два подхода к определению последовательности фонем:

- алгоритмы на основании цепи Маркова: на вход алгоритма поступают фонемы с вероятностями, из которых составляется последовательность;
- динамическое программирование: в каждый момент времени определяется предполагаемая последовательность фонем от начала до текущего момента времени.

### Синтез речи

Обратный процесс распознаванию речи, не осложненный обработкой шумов и акцентов, выступает синтез речи (рис. 3). Основной задачей при синтезе речи является превращение текста в звуковую волну, наиболее приближенную к человеческому голосу.



Рис. 3. Алгоритм синтеза речи

Первый этап – лингвистическая обработка. Она предусматривает, во-первых, исправление грамматических и синтаксических ошибок (для постановки правильной интонации), во-вторых, представление текста в едином формате (расшифровка числовых последовательностей, обработка аббревиатур и иноязычных вставок). Второй, третий и четвертый этапы синтеза речи предусматривают деление текста на составляющие до тех пор, пока каждую часть нельзя описать числами. Полученные неделимые части (в зависимости от используемой модели имеют разную размерность), называемые фреймы, описывают параметрами, из которых затем создают акустическую модель. Благодаря акустической модели генерируют звуковые волны, воспринимаемые человеком как речь.

В настоящее время выделяют две основные модели синтеза речи (таблица 1).

*Таблица 1. Модели синтеза речи*

Название модели синтеза	Описание	Достоинства	Недостатки
Компилятивный синтез	Конкатенации (склейки) ранее записанных диктором звуковых фрагментов	<ul style="list-style-type: none"> <li>• высокое качество звука;</li> <li>• естественность звучания;</li> <li>• высокая скорость синтеза</li> </ul>	<ul style="list-style-type: none"> <li>• низкая вариативность;</li> <li>• большой объем данных для обучения (около 20 часов чистой речи)</li> </ul>
Параметрический синтез	Построение модели, описывающей акустический свойства аудиосигнала для данного текста	<ul style="list-style-type: none"> <li>• возможность генерации разнообразной интонации;</li> <li>• меньшее количество объема данных по сравнению с компилятивным синтезом</li> </ul>	<ul style="list-style-type: none"> <li>• низкая скорость работы;</li> <li>• большая вычислительная сложность (из-за применения глубокого обучения)</li> </ul>

В рамках рассматриваемой задачи используется модель параметрического синтеза, поскольку важна гибкость произношения, при этом нет необходимости синтезировать большой объем текста.

### **Меры сходства текстов**

Одной из задач при написании программы выступает поиск команд, соответствующих голосовому запросу. Поэтому голосовой запрос, преобразованный в текст, подвергается дальнейшему анализу для нахождения подобия.

Прежде чем приступить к расчету коэффициента схожести текст подвергается предварительной обработки:

- замена множественных пробелов на одинарные;
- разбиение на токены (разделенные пробелами и знаками препинания символы);

- удаление знаков препинания;
- перевод в нижний регистр;
- нормализация (стемминг-выделения основ слова, лемматизация-приведение слов к начальной форме);
- подсчет вхождений.

После предварительной обработки слова представляют вектора, над которыми можно производить арифметические действия.

Существует несколько метрик сходства. Самыми популярными для определения схожести коротких предложений выступают косинусная мера сходства и мера Жаккара. Обе формулы оперируют такими величинами как  $\vec{v}$  – вектор слова из первого текста, а  $\vec{w}$  – вектор того же слова из второго текста.

Косинусная мера сходства:

$$sim_{cosine}(\vec{v}, \vec{w}) = \frac{\vec{v} \cdot \vec{w}}{|\vec{v}| \cdot |\vec{w}|} = \frac{\sum_{i=1}^N v_i \cdot w_i}{\sqrt{\sum_{i=1}^N v_i^2} \cdot \sqrt{\sum_{i=1}^N w_i^2}}$$

Мера Жаккара:

$$sim_{Jaccard}(\vec{v}, \vec{w}) = \frac{\sum_{i=1}^N \min(v_i, w_i)}{\sum_{i=1}^N \max(v_i, w_i)}$$

Для определения важности слов, необходимо учитывать количество вхождений слова в запрос, а значит косинусная мера сходства является наиболее подходящим вариантом для решения этой задачи, поскольку в мере Жаккара учитываются лишь максимальные и минимальные значения вхождения слова.

### **Реализация**

При создании программного обеспечения был проведен анализ механизмов распознавания и синтеза речи, выбраны оптимальные средства разработки. На основании библиотек было разработано приложение на языке Python, выполняющее роль голосового ассистента.

Запуск приложения происходит с платформы 1С:Предприятие (рис. 4). Если exe-файл не найден на локальном компьютере, приложение загружается с сервера и запускается с указанными параметрами подключения.

Стоит заметить, что данное приложение обладает универсальностью, а именно возможностью подключения к любой информационной базе 1С:Предприятие. Нет необходимости произношения дословных команд, поскольку вычисляется мера сходства, на основании которой происходит выбор. Базу доступных команд можно пополнять любыми запросами.

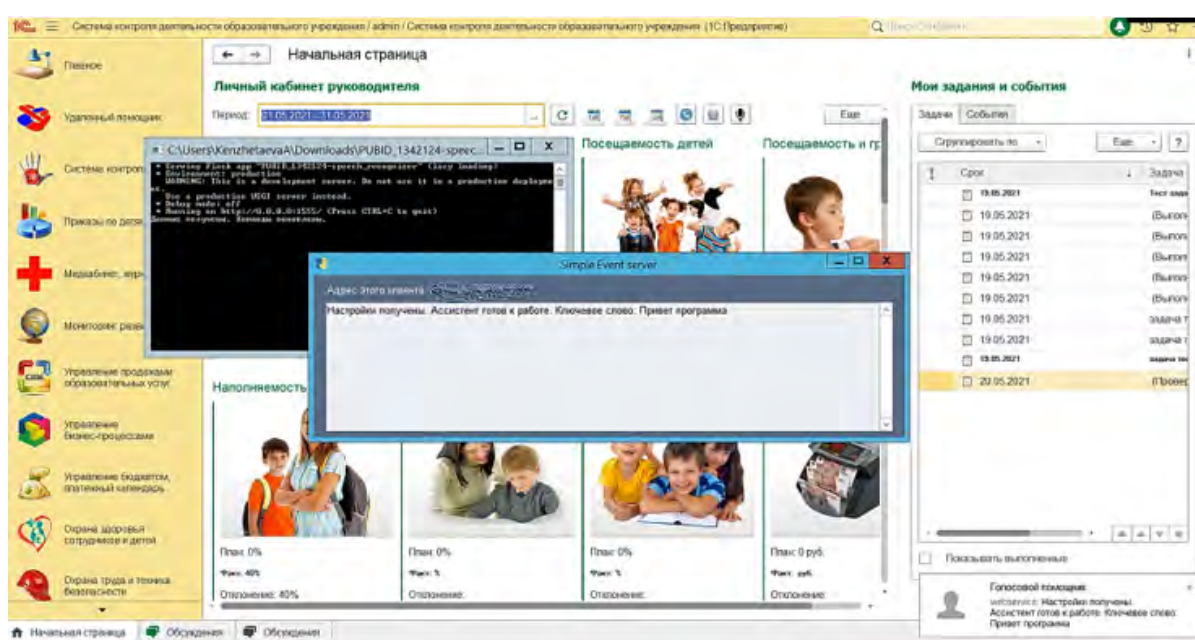


Рис. 4. Запуск приложения

Обращение к базе и выполнение команд начинается после произнесения активационной фразы, которую пользователь может указать самостоятельно в настройках. Все произнесенные слова отображаются в окне приложения (рис. 5). Также диалог с пользователем виден через систему взаимодействия, непосредственно на платформе 1С: Предприятие (рис. 6).



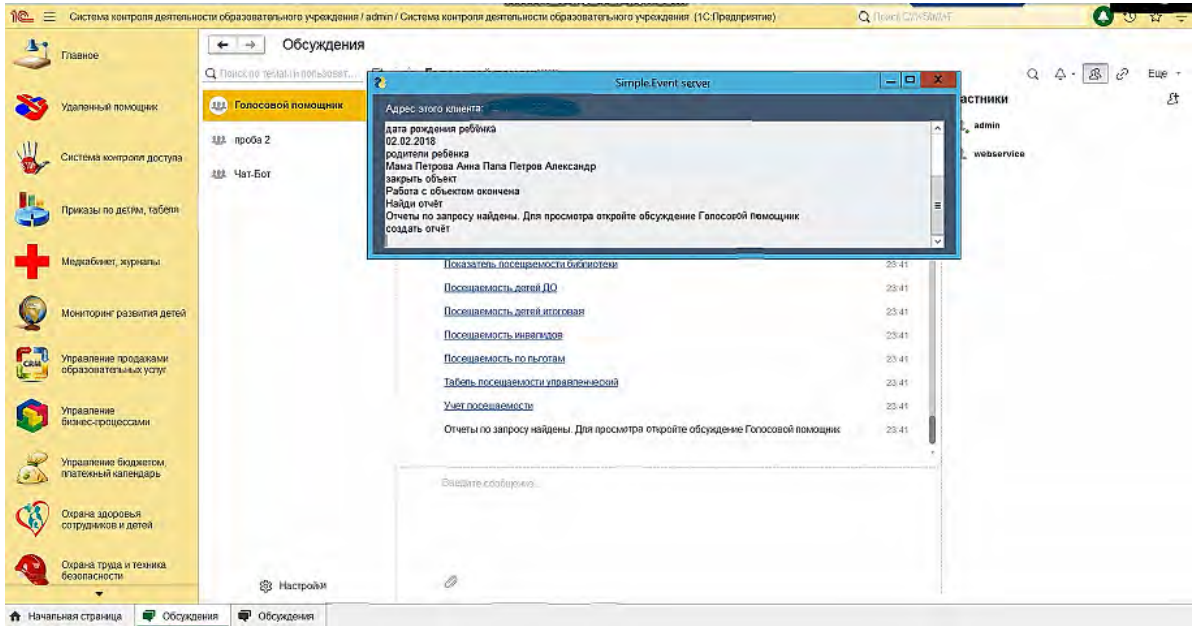


Рис. 5. Отображение диалога в приложении

На данный момент программное приложение справляется с рядом задач:

- получение информации из базы;
- отправка ссылок на необходимые объекты;
- создание напоминаний, задач;
- формирование отчетов, с указанием необходимых параметров;
- открытие необходимых форм.

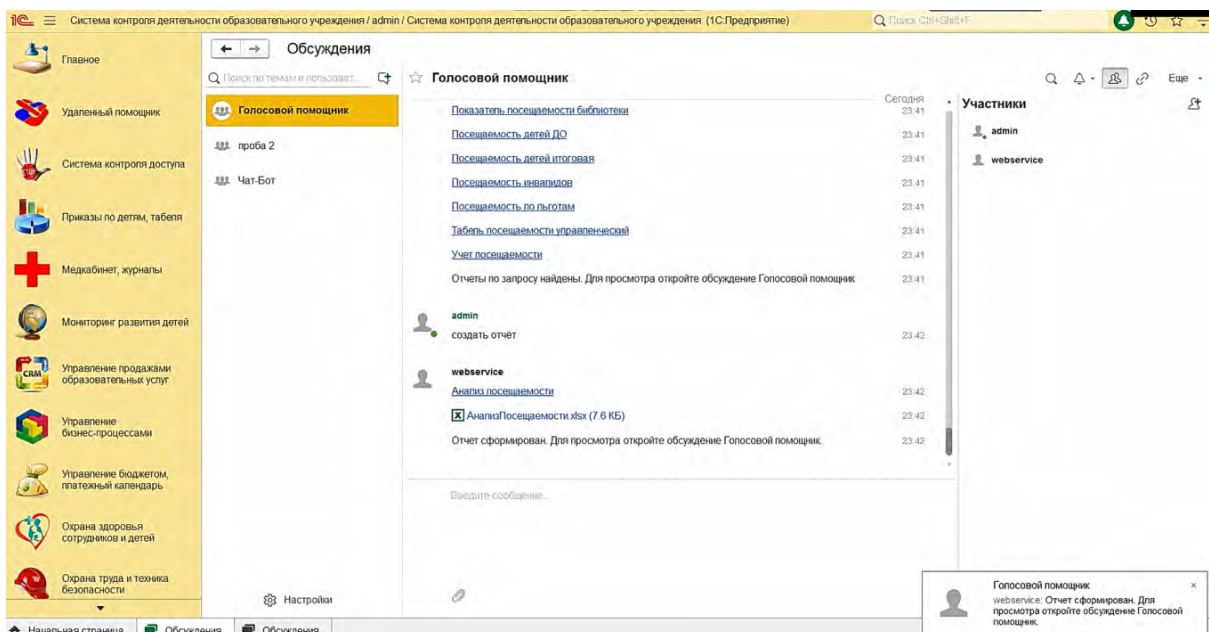


Рис. 6. Отображение диалога в обсуждении

## **Заключение**

В результате выполнения поставленных задач было разработано программное приложение «Голосовой помощник для платформы 1С: Предприятие». Данное приложение полностью справляется с возложенными на нее обязанностями, а именно получением необходимой информации в кратчайшие сроки, выполнение рутинной работы пользователей, а также управление платформой голосом.

При создании программного приложения были выполнены ряд задач:

- разработан алгоритм работы голосовых помощников;
- проанализированы методы распознавания и синтеза голоса;
- проведен сравнительный анализ мер сходства для текста.

Направления дальнейшей работы:

- реализация интерактивного создания объектов;
- обучение системы на отвлеченный диалог.

## **СПИСОК ЛИТЕРАТУРЫ**

1. Обзор технологий синтеза речи: [Электронный ресурс]. URL: <https://habr.com/ru/company/tinkoff/blog/474782/> (Дата обращения: 18.05.2021).
2. Синтез речи: учебное пособие / С.В. Рыбин. – СПб.: Университет ИТМО, 2014.
3. Электронный научно-практический журнал «Современные научные исследования и инновации»: [Электронный ресурс]. URL: <https://web.snauka.ru/> (Дата обращения: 18.05.2021).
4. Рассчитать сходство – самые важные метрики в двух словах: [Электронный ресурс]. URL: <https://www.machinelearningmastery.ru/calculate-similarity-the-most-relevant-metrics-in-a-nutshell-9a43564f533e/> (Дата обращения: 18.05.2021).

5. SpeechKit – речевые технологии Яндекса: [Электронный ресурс].  
URL: [https://yandex.ru/company/technologies/speech\\_technologies/](https://yandex.ru/company/technologies/speech_technologies/) (Дата обращения: 17.05.2021).

6. Текстонезависимая идентификация по голосу: [Электронный ресурс].  
URL: <https://m.habr.com/ru/post/336516/> (Дата обращения: 19.05.2021).