
КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ И АВТОМАТИЗАЦИЯ ПРОЕКТИРОВАНИЯ/COMPUTER MODELING AND DESIGN AUTOMATION

DOI: <https://doi.org/10.60797/IRJ.2026.169.69> EDN: DAWNDR**ГОЛОСОВОЕ УПРАВЛЕНИЕ ДЛЯ СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ**

Научная статья

Гизатуллин З.М.^{1,*}, Нуртдинов Р.С.²¹ORCID : 0000-0003-0571-5593;^{1,2} Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ, Казань, Российская Федерация

* Корреспондирующий автор (gzm_zinnur[at]mail.ru)

Предложена: 06.03.2026; Принята: 06.07.2026; Опубликовано: 17.07.2026

Аннотация

Работа посвящена разработке голосового интерфейса для системы автоматизации проектирования (САПР) на примере программы FreeCAD. Рассматриваются современные технологии автоматического распознавания речи, включая глубокие нейросетевые модели, и анализируется возможность их использования в задаче голосового управления САПР. На основе анализа предложена эффективная комбинация нейросетевых и программных архитектур для голосового интерфейса САПР, обеспечивающая надежное распознавание русской речи, в том числе в условиях шумов, и выполнение соответствующих команд в режиме реального времени. Разработано приложение и получено свидетельство о регистрации прототипа системы голосового интерфейса САПР, позволяющей повысить доступность, эргономику и эффективность инженерного проектирования.

Ключевые слова: система автоматизированного проектирования, голосовое управление, FreeCAD, распознавание речи, нейронная сеть.

VOICE CONTROL FOR A COMPUTER-AIDED DESIGN SYSTEM

Research article

Gizatullin Z.M.^{1,*}, Nurtudinov R.S.²¹ORCID : 0000-0003-0571-5593;^{1,2} Kazan National Research Technical University named after A.N. Tupolev-KAI, Kazan, Russian Federation

* Corresponding author (gzm_zinnur[at]mail.ru)

Suggested: 06.03.2026; Accepted: 06.07.2026; Published: 17.07.2026

Abstract

The paper focuses on the development of a voice interface for a computer-aided design (CAD) system on the example of the FreeCAD programme. Modern automatic speech recognition technologies, including deep neural network models, are examined, and the possibility of their use in the task of voice control for CAD systems is analysed. Based on this analysis, an effective combination of neural network and software architectures is suggested for a CAD voice interface, ensuring reliable recognition of Russian speech, including in noisy environments, and the execution of corresponding commands in real time. An application has been developed and a registration certificate obtained for a prototype CAD voice interface system, which improves the accessibility, ergonomics and efficiency of engineering design.

Keywords: computer-aided design system, voice control, FreeCAD, speech recognition, neural network.

Введение

В настоящее время системы автоматизированного проектирования (САПР) занимают важное место во многих отраслях промышленности — от машиностроения и архитектуры до электроники и промышленного дизайна. Развитие САПР привело к появлению мощных программных комплексов с богатой функциональностью, но вместе с тем и к усложнению пользовательских интерфейсов. Исследования показывают, что навигация в многофункциональных интерфейсах САПР может занимать значительное время, особенно у начинающих пользователей, и непонимание элементов графического интерфейса может стать серьезным барьером для эффективной работы [1]. Таким образом, данные факторы замедляют их освоение и могут снижать общую продуктивность проектирования [2].

В связи с этим особую актуальность приобретает внедрение голосовых интерфейсов в САПР, которые призваны упростить и ускорить взаимодействие пользователя с системой. Голосовое управление позволяет передавать команды естественным способом — голосом, без необходимости отвлекаться на поиск нужной кнопки или ввода с клавиатуры. Интуитивно понятный голосовой ввод способен уменьшить когнитивную нагрузку на проектировщика и сократить время на выполнение операций. Отмечается, что интеграция интуитивных способов ввода снижает барьеры использования сложных систем и делает их более доступными для широкого круга специалистов [3].

Цель данной работы — создание прототипа интерфейса голосового управления для САПР на примере FreeCAD. FreeCAD — популярная САПР с открытым исходным кодом, имеет встроенный Python-интерпретатор и интерфейс scripting API. Из консоли Python или из пользовательских скриптов можно управлять FreeCAD и выполнять все

операции, доступные через его графический интерфейс. Это значит, что результат распознавания голоса можно напрямую преобразовывать в вызовы функций FreeCAD через его Python API.

Научная новизна работы заключается в построении эффективной комбинации нейросетевых и программных архитектур интерфейса голосового управления САПР, обеспечивающего надежное распознавание русской речи, в том числе в условиях шумов, и выполнение соответствующих команд в режиме реального времени.

Методы и принципы исследования

Современные системы голосового управления состоят из подсистем распознавания речи и последующего семантического разбора команд. К ключевым открытым технологиям распознавания речи (ASR-движкам) относятся Vosk, OpenAI Whisper и Mozilla DeepSpeech/Coqui, а также коммерческие облачные сервисы (Google Speech-to-Text, Yandex SpeechKit, Microsoft Azure Speech и др.).

Vosk — легковесный оффлайн-фреймворк на базе Kaldi, поддерживающий более 20 языков, работающий даже на слабых устройствах и без интернет-соединения. Он имеет малый размер языковых моделей и прост в установке, поддерживает потоковое распознавание и предлагает привязки к разным языкам программирования [4].

Компания Fraunhofer IDMT представила решение для надёжного управления производственным оборудованием с помощью голоса даже в шумной обстановке, также существуют коммерческие системы вроде Voice2CAD, которые задействуют технологии распознавания речи Microsoft для выполнения команд в AutoCAD, Revit, SolidWorks и других пакетах — фактически озвучивая стандартные горячие клавиши и macros CAD-систем. В работе [5] описана система VOICE CAD для трёхмерного моделирования голосовыми командами, благодаря которой время выполнения типовых операций сократилось на ~40% по сравнению с традиционным интерфейсом, а точность распознавания команд достигла 92%. Пользователи без опыта работы с CAD смогли приступить к работе с таким голосовым интерфейсом всего за 15 минут обучения, демонстрируя 90%-й уровень удовлетворённости системой.

OpenAI Whisper — нейросетевая модель с открытым кодом, созданная на основе 680 тыс. часов многоязычных аудиоданных. Это самый популярный инструмент транскрибации, обеспечивающий очень высокое качество распознавания. Whisper поддерживает десять языков, обладает повышенной устойчивостью к шумам, акцентам и техническому жаргону, однако модели Whisper требуют до 10 ГБ GPU-памяти и работают партиями по 30-секундным фрагментам, что увеличивает задержку [6].

Mozilla DeepSpeech — открытый движок ASR, основанный на движке Baidu Deep Speech. Он обеспечивает потоковое распознавание речи и может работать в реальном времени на GPU и CPU, так же у него есть поддержка TensorFlow Lite для мобильных устройств. DeepSpeech позволяет обучать собственные языковые модели, но набор поддерживаемых языков ограничен, а для подготовки исходных данных требует большое количество времени. В реальных программах DeepSpeech подходит для относительно небольших словарей команд на английском языке и требует постобработки результата [4].

Помимо оффлайн-решений, применяются так же облачные сервисы распознавания. Google Speech-to-Text API поддерживает более 125 языков и диалектов и демонстрирует очень высокую точность благодаря большим нейросетевым моделям. Сервис легко интегрируется через REST/gRPC API, может работать в потоковом режиме и адаптироваться под специфическую лексику [7]. Недостатки облачных решений — необходимость постоянного интернет-соединения, оплата по времени обработки и потенциальные риски конфиденциальности. Для русского языка используют Yandex SpeechKit, облачный API с хорошей поддержкой фоновой речи и словарей русского языка, Microsoft Azure Speech-to-Text и IBM Watson.

После конвертации речи в текст нужно выделить смысл команды. Для этого используются NLP/NLU-библиотеки и платформы. Rasa Open Source — фреймворк на Python для построения диалоговых систем. Rasa позволяет создавать контекстно-зависимые ассистенты и легко подключается к различным каналам [8]. Snips NLU — легковесная Python-библиотека для извлечения структурированной информации из естественного языка (определения интента и параметров команды) [9]. NLTK (Natural Language Toolkit) — классический набор Python-библиотек для NLP (токенизация, стемминг, теггинг, парсинг, семантический анализ и др.) [10].

Основные характеристики различных подходов к распознаванию речи приведены в таблице 1.

Таблица 1 - Сравнение подходов к распознаванию голосовых команд по ключевым параметрам

DOI: <https://doi.org/10.60797/IRJ.2026.169.69.1>

Подход/Платформа	Оффлайн	Словарь	Устойчивость к шуму	Гибкость лексики
Облачные сервисы	Нет	Очень большой	Очень высокая	Низкая (фиксированная модель)
Vosk (Kaldi, TDNN-LSTM)	Да	Большой (динамический)	Высокая	Высокая (расширяемый словарь)
Грамматические шаблоны (WFST)	Да	Ограниченный	Средняя	Средняя (фиксированные шаблоны)

При реализации современных систем голосового управления одним из ключевых направлений является обеспечение шумоустойчивости и малой задержки системы [11]. Для повышения устойчивости исследуются методы предобработки и аугментации: добавление шума на этапе обучения, спектральные преобразования и генеративные модели могут улучшать распознавание в шумах и помехах [12]. Вместе с тем голосовой интерфейс зачастую должен работать в реальном времени. Это накладывает ограничения на архитектуру: требуется потоковая обработка и малая задержка распознавания. В этой связи рассматриваются модели без автокорреляции и с жесткими временными ограничениями.

Основные результаты

Для реализации голосового управления САПР разработан алгоритм, последовательно выполняющий преобразование устной команды пользователя в соответствующее действие в среде FreeCAD. Алгоритм состоит из ряда этапов от захвата звукового сигнала до исполнения команды в САПР. Ниже приведены основные стадии этого процесса:

1. Ввод голосовой команды. Пользователь произносит команду, которая улавливается микрофоном и преобразуется в цифровой аудиосигнал. На этом этапе осуществляется предварительная обработка, оцифровка речи и выделение MFCC-признаков (набор числовых признаков, описывающих спектральные свойства речевого сигнала в нелинейной мел-шкале частот) для подготовки данных к распознаванию [13].

2. Распознавание речи. На следующем этапе происходит непосредственное распознавание речи — преобразование последовательности признаков в текстовую последовательность. Для решения данной задачи используется система Vosk, которая базируется на гибридной архитектуре DNN-HMM (глубокая нейронная сеть и скрытая марковская модель). Акустическая модель Vosk представляет собой глубокую нейронную сеть, состоящую из комбинации TDNN (нейронная сеть с временной задержкой) и LSTM (сеть долгой краткосрочной памяти) слоёв [14]. На вход данной системы подаются MFCC-признаки, а на выходе получаем текстовую транскрипцию голосовой команды.

3. Интерпретация команды. На этапе интерпретации команды в разработанном прототипе используется словарный алгоритм прямого сопоставления. Словарь команд хранится во внешнем JSON-файле `commands.json`; для каждой команды задаются основное имя, список синонимов и программный код действия. После распознавания речи полученная строка последовательно сравнивается с записями словаря: сначала с основным именем команды, затем с её синонимами. При полном совпадении выбирается соответствующая команда и передаётся на исполнение в FreeCAD. Таким образом, вариативность пользовательских формулировок обеспечивается за счёт заранее добавляемых синонимов, а порог срабатывания в базовой реализации является бинарным: команда выполняется только при точном совпадении распознанной строки с одной из допустимых формулировок. Настройка интерпретации осуществляется редактированием JSON-словаря и списка синонимов через интерфейс программы. Такой подход обеспечивает простоту и предсказуемость работы, однако ограничивает гибкость системы, поскольку она не учитывает смысл свободно сформулированного запроса и зависит от заранее заданного словаря.

4. Подтверждение команды пользователем. Перед выполнением действия система предусматривает механизм подтверждения, особенно актуальный в случаях неуверенного распознавания или неоднозначности команды. Если полученная на предыдущем этапе команда имеет низкую степень уверенности распознавания либо найдены несколько похожих вариантов, модуль интерфейса запрашивает у пользователя подтверждение или уточнение. В случае подтверждения пользовательским вводом команда считается корректной и передаётся на исполнение. Если же пользователь сообщает об ошибке или повторяет команду (при запросе повтора), алгоритм возвращается к этапу распознавания речи.

5. Передача команды в документ САПР. Определив необходимую команду, система передаёт её на выполнение в среду FreeCAD с помощью встроенного Python API. FreeCAD обладает встроенным интерпретатором Python и предоставляет программный доступ ко всем основным функциям системы. Во всех случаях цель заключается в том, чтобы активировать соответствующую функцию FreeCAD так, словно пользователь выбрал её через графический интерфейс, но посредством программного вызова.

6. Обратная связь и отображение результата. На финальном этапе пользователь получает подтверждение о выполнении команды и её результат. Если команда успешно исполнена, то результат видно непосредственно в среде — например, на модели появляется созданный объект, изменяется геометрия или выполняется другое ожидаемое действие.

Таким образом, формализованная постановка задачи выглядит следующим образом: преобразовать акустический сигнал A в текстовую последовательность: $R = \text{SpeechToText}(A)$; интерпретировать текстовую последовательность в команду: $R \rightarrow C_i$; осуществить функциональное отображение команды C_i на конкретную исполняемую функцию f из программного интерфейса САПР: $C_i \rightarrow f \in \text{API}$. Ограничение на время распознавания голосовой команды для комфортной работы менее 500 мс. При этом необходимо максимизировать функционал качества $Q = N_c / N_t$, где N_c — количество корректно распознанных и выполненных инженерных команд; N_t — общее количество голосовых команд.

Ограничения текущей реализации: система использует ограниченный словарь команд и синонимов, поэтому корректная интерпретация возможна только для заранее заданных формулировок; ограничение на допустимую длину команды: предполагается, что голосовая команда представляет собой относительно короткую фразу (как правило, одно слово или команда из нескольких слов).

UML диаграмма компонентов разработанной системы представлена на рисунке 1.

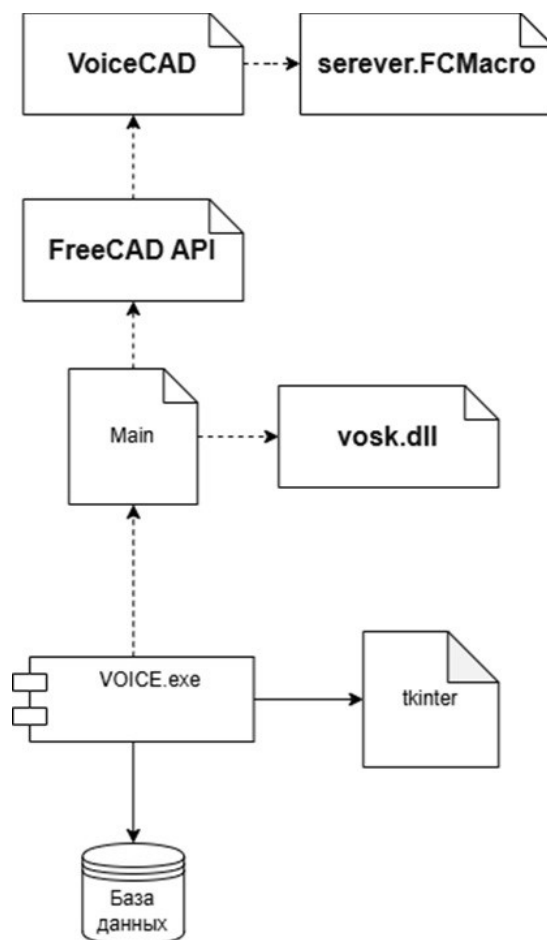


Рисунок 1 - UML диаграмма компонентов
DOI: <https://doi.org/10.60797/IRJ.2026.169.69.2>

Основным компонентом представленной диаграммы является исполняемый файл VOICE.exe. Данный компонент реализует весь основной функционал голосового управления и представляет собой приложение, запускаемое на рабочей станции пользователя. Взаимодействие с базой данных осуществляется напрямую из VOICE.exe. Компонент tkinter реализует графический пользовательский интерфейс приложения VOICE.exe. Через этот интерфейс пользователь может управлять списком команд, просматривать историю команд, изменять параметры распознавания речи и настраивать макросы. Взаимодействие между VOICE.exe и tkinter реализовано программно – через вызовы методов библиотеки tkinter для создания и обработки окон, диалогов и пользовательских событий.

Модуль Main отвечает за организацию внутренней логики работы приложения, запуск и контроль всех процессов в VOICE.exe, а также интеграцию внешних библиотек и модулей. Main инициализирует взаимодействие с модулем распознавания речи и с компонентом FreeCAD API.

Для распознавания голосовых команд используется внешняя библиотека vosk.dll (Vosk Speech Recognition). Связь между Main и vosk.dll реализована через вызовы API библиотеки — Main инициирует процесс захвата аудиосигнала, передачи его на распознавание и получения текстового результата.

Для передачи команд непосредственно в FreeCAD используется FreeCAD API. Это компонент, реализующий взаимодействие с FreeCAD через его открытый Python API. VOICE.exe формирует текст команды и передаёт его в FreeCAD API, где команда исполняется непосредственно в CAD-среде.

В составе FreeCAD отдельным компонентом выделен макрос sever.FCMacro, который обеспечивает приём и обработку входящих команд из внешних систем. Связь с этим макросом осуществляется по механизму inter-process communication (обычно — через сокеты или внутренний сервер макроса в FreeCAD), что позволяет получать команды, отправленные из VOICE.exe или других приложений, и обрабатывать их в реальном времени.

Пользовательский интерфейс программы представляет собой набор пиктограмм (иконок), предназначенных для доступа к разнообразным функциям программ, кнопки для выполнения различных операций, а также окна, на которых демонстрируются данные во время взаимодействия с программным продуктом.

Для обеспечения удобства взаимодействия пользователя с программой была разработана схема навигации, отражающая логику переходов между основными окнами и модулями системы (рисунок 2).

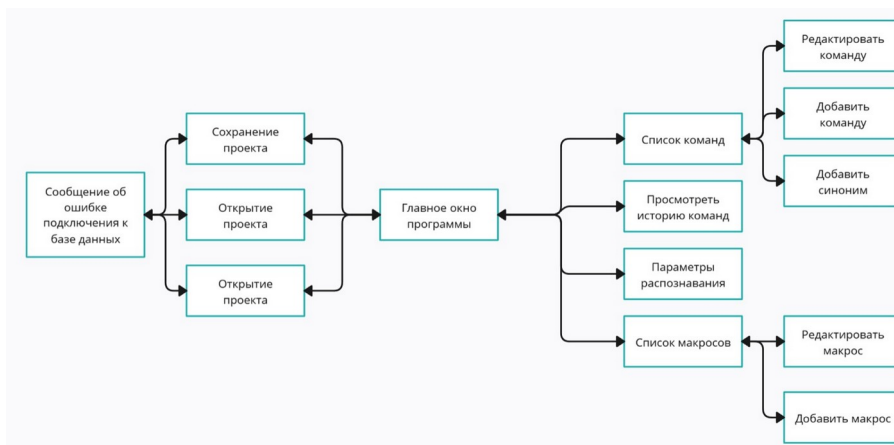


Рисунок 2 - Схема навигации между основными окнами и модулями системы

DOI: <https://doi.org/10.60797/IRJ.2026.169.69.3>

Интерфейс модуля «параметры распознавания» демонстрируют функциональное окно, предназначенные для управления параметрами, языком распознавания. Данный интерфейс представлен на рисунке 3.

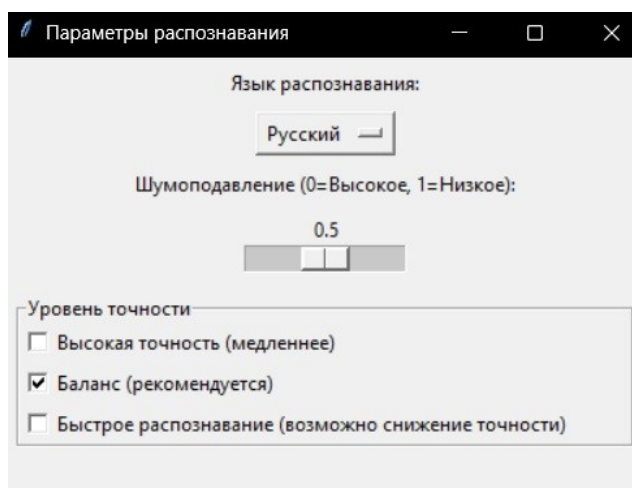


Рисунок 3 - Окно параметров распознавания речи

DOI: <https://doi.org/10.60797/IRJ.2026.169.69.4>

Обсуждение

Пример результат работы голосового интерфейса в системе FreeCAD представлена на рисунке 4.

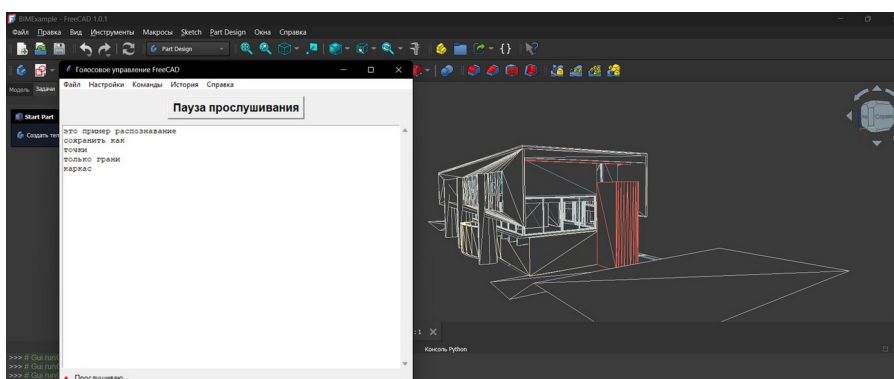


Рисунок 4 - Главное окно голосового интерфейса и результат исполнения команды «каркас» в FreeCAD

DOI: <https://doi.org/10.60797/IRJ.2026.169.69.5>

В разработанном прототипе автоматизированы функции, связанные как с вызовом команд FreeCAD, так и с управлением словарём голосовых команд. Это, например, команды «сохранить как», команды переключения режимов отображения модели «точки», «только грани», «каркас», «как есть», а также команды управления видом «повернуть влево», «повернуть вправо». Система поддерживает добавление, редактирование и удаление команд, расширение набора синонимов через интерфейс программы и ведение истории выполненных действий.

Для оценки эффективности разработанного прототипа была проведена предварительная экспериментальная апробация на наборе из 10 типовых голосовых команд FreeCAD, состоящих из одного или двух слов. Команды в наборе: «сохранить», «сохранить как», «отменить», «повторить», «каркас», «точки», «только грани», «как есть», «повернуть влево», «повернуть вправо». В эксперименте участвовали 3 диктора: два студента, юноша и девушка, а также один преподаватель мужского пола. Каждая команда произносилась каждым диктором по 10 раз в двух акустических условиях таких как:

- 1) в комнате преподавателя при отсутствии фоновой речи и выраженных источников шума;
- 2) в той же комнате при работающем напольном вентиляторе на расстоянии около 2 м, с заявленным уровнем шума не более 46 дБ.

Для записи использовался встроенный микрофон ноутбука Lenovo Legion 5.

Таким образом, на одного диктора приходилось 100 произнесений команд в одном акустическом условии и 200 произнесений в двух условиях. Общий объём выборки составил 600 тестовых произнесений (таблица 2).

Таблица 2 - Сводные результаты распознавания команд

DOI: <https://doi.org/10.60797/IRJ.2026.169.69.6>

Условие тестирования	Количество произнесений	Корректно распознано	Неверно выбрана другая команда	Команда не распознана	Точность, %
1. Без выраженного фонового шума	300	253	31	16	84,3
2. При работе вентилятора	300	238	38	24	79,3
Итого	600	491	69	40	81,8

При обработке каждого произнесения фиксировался один из трёх взаимоисключающих исходов: корректное распознавание целевой команды; неверный выбор другой команды из словаря; отсутствие распознавания команды.

Под неверным выбором другой команды понималась ситуация, когда вместо произнесённой команды система активировала другую команду из заданного набора. Под нераспознанной командой понималась ситуация, когда команда не была сопоставлена ни с одним элементом словаря и действие не выполнялось.

По результатам эксперимента средняя точность распознавания составила 81,8%. Для данной выборки 95%-й доверительный интервал составляет $\pm 3,1\%$, то есть итоговая оценка точности составляет $81,8 \pm 3,1\%$. Снижение точности во втором акустическом условии связано с наличием дополнительного фонового шума, что увеличило как количество ошибок замены команды, так и число нераспознанных команд. Полученный результат соответствует общим тенденциям, отмечаемым в литературе для систем распознавания команд с ограниченным словарём, где точность существенно зависит от предметной настройки словаря, акустических условий, качества микрофона и уровня фонового шума. Для Vosk-подобных решений повышение точности достигается, в частности, за счёт адаптации словаря и языковой модели под конкретный домен команд.

На разработанное приложение получено свидетельство о государственной регистрации программы для ЭВМ №2025689922 РФ «Голосовой интерфейс для программы FreeCAD».

Заключение

В работе проведен анализ решений в области голосового управления и предложена эффективная комбинация нейросетевых и программных архитектур голосового интерфейса для системы САПР, обеспечивающее надежное распознавание русской речи и выполнение соответствующих команд в режиме реального времени. На основании предложенных архитектур создан прототип интерфейса голосового управления для FreeCAD. На разработанное приложение получено свидетельство о государственной регистрации программы для ЭВМ «Голосовой интерфейс для программы FreeCAD».

В принципе аналогичная интеграция возможна и с другими open-source CAD-системами, однако её сложность определяется доступностью программного интерфейса. Для OpenSCAD интеграция представляется возможной через генерацию .scad-сценариев и запуск в командной строке, поскольку система ориентирована на параметрическое текстовое описание моделей и работу в command-line environment. Для LibreCAD интеграция также принципиально возможна, но требует отдельного адаптера: в актуальной ветке LibreCAD 2.x документированы command line и console tool, тогда как в LibreCAD 3 заявлена возможность Lua-скриптинга. При этом перенос прототипа на LibreCAD и OpenSCAD потребует переработки словаря команд и исполнительного слоя, поскольку LibreCAD ориентирован на 2D-черчение, а OpenSCAD — на сценарное 3D-моделирование, то есть семантика команд и набор доступных операций в них отличаются от FreeCAD.

Перспективой развития предлагаемого подхода голосового управления является переход от словарного голосового интерфейса к интеллектуальной голосовой среде, в которой подключаемый модуль искусственного интеллекта, например YandexGPT или DeepSeek, будет интерпретировать пользовательский запрос с учётом контекста и сопоставлять его с командами не только по точному совпадению, но и по смыслу. Дополнительно предполагается расширение системы до многосредового режима работы, при котором пользователь сможет выбирать приложение для голосового управления. Помимо FreeCAD, в единую оболочку может быть включён уже реализованный модуль голосового управления Microsoft Word, а в дальнейшем — и другие прикладные программы.

Конфликт интересов

Не указан.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Conflict of Interest

None declared.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Зинченко Ю.В. Обзор современных систем автоматизированного проектирования / Ю.В. Зинченко, А.А. Голобородько // Потенциал современной науки. — 2016. — 4. — С. 68–71.
2. Szewczyk J. Difficulties with the novices' comprehension of the computer-aided design (CAD) interface: understanding visual representations of CAD tools / J. Szewczyk // Journal of Engineering Design. — 2003. — 14. — P. 169–185. — DOI: 10.1080/0954482031000091491
3. Rehman S.U. Advancing BIM and game engine integration in the AEC industry: innovations, challenges, and future directions / S.U. Rehman, I. Kim, K.-E. Hwang // Journal of Computational Design and Engineering. — 2025. — 4. — P. 26–54. — DOI: 10.1093/jcde/qwaf030
4. Обзор популярных open-source инструментов для транскрипции видео/аудио. — 2025. — URL: <https://officelife.media/article/colleagues-say/60503-obzor-populyarnykh-open-source-instrumentov-dlya-transkribatsii-video-audio> (дата обращения: 24.02.2026).
5. Sawalkar V. VOICECAD: revolutionizing 3D modeling through voice commands / V. Sawalkar, A. Patare, A. Gorlawar et al. // International Journal of Advanced Research in Science, Communication and Technology. — 2024. — 4. — P. 639–648. — DOI: 10.48175/568
6. Whisper by OpenAI. — URL: <https://openai.com/index/whisper/> (accessed: 24.02.2026).
7. Google Cloud Speech-to-Text. — URL: <https://cloud.google.com/speech-to-text> (accessed: 24.02.2026).
8. RasaHQ/rasa. — URL: <https://github.com/RasaHQ/rasa> (accessed: 24.02.2026).
9. Snips-nlu. — URL: <https://github.com/snipsco/snips-nlu> (accessed: 24.02.2026).
10. NLTK. — URL: <https://www.nltk.org/> (accessed: 24.02.2026).
11. Dua M. Noise robust automatic speech recognition: review and analysis / M. Dua, Akanksha, S. Dua // International Journal of Speech Technology. — 2023. — 2. — P. 475–519.
12. Khusnutdinov A.N. The Sound Pressure Level Meter / A.N. Khusnutdinov, M.G. Nuriev. // 2022 International Russian Automation Conference (RusAutoCon); — Sochi: IEEE, 2022. — P. 63–68. doi: 10.1109/RusAutoCon54946.2022.9896267
13. Овчинников Е.Л. Частотное восприятие звука: Высота тонов в технике: мел-шкала / Е.Л. Овчинников // Наука и образование в современном мире. — 2016. — 11. — С. 39–42.
14. Huang C.L. Exploring effective data augmentation with TDNN-LSTM neural network embedding for speaker recognition / C.L. Huang. // 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU); — 2019: IEEE, 2019. — P. 291–295.

Список литературы на английском языке / References in English

1. Zinchenko Yu.V. Obzor sovremenny'x sistem avtomatizirovannogo proektirovaniya [Review of modern computer-aided design systems] / Yu.V. Zinchenko, A.A. Goloborod'ko // The potential of modern science. — 2016. — 4. — P. 68–71. [in Russian]
2. Szewczyk J. Difficulties with the novices' comprehension of the computer-aided design (CAD) interface: understanding visual representations of CAD tools / J. Szewczyk // Journal of Engineering Design. — 2003. — 14. — P. 169–185. — DOI: 10.1080/0954482031000091491
3. Rehman S.U. Advancing BIM and game engine integration in the AEC industry: innovations, challenges, and future directions / S.U. Rehman, I. Kim, K.-E. Hwang // Journal of Computational Design and Engineering. — 2025. — 4. — P. 26–54. — DOI: 10.1093/jcde/qwaf030
4. Obzor populyarnykh open-source instrumentov dlya transkribatsii video/audio [Review of popular open-source tools for video/audio transcription]. — 2025. — URL: <https://officelife.media/article/colleagues-say/60503-obzor-populyarnykh-open-source-instrumentov-dlya-transkribatsii-video-audio> (accessed: 24.02.2026). [in Russian]
5. Sawalkar V. VOICECAD: revolutionizing 3D modeling through voice commands / V. Sawalkar, A. Patare, A. Gorlawar et al. // International Journal of Advanced Research in Science, Communication and Technology. — 2024. — 4. — P. 639–648. — DOI: 10.48175/568
6. Whisper by OpenAI. — URL: <https://openai.com/index/whisper/> (accessed: 24.02.2026).



7. Google Cloud Speech-to-Text. — URL: <https://cloud.google.com/speech-to-text> (accessed: 24.02.2026).
8. RasaHQ/rasa. — URL: <https://github.com/RasaHQ/rasa> (accessed: 24.02.2026).
9. Snips-nlu. — URL: <https://github.com/snipsco/snips-nlu> (accessed: 24.02.2026).
10. NLTK. — URL: <https://www.nltk.org/> (accessed: 24.02.2026).
11. Dua M. Noise robust automatic speech recognition: review and analysis / M. Dua, , Akanksha, S. Dua // International Journal of Speech Technology. — 2023. — 2. — P. 475–519.
12. Khusnutdinov A.N. The Sound Pressure Level Meter / A.N. Khusnutdinov, M.G. Nuriev. // 2022 International Russian Automation Conference (RusAutoCon); — Sochi: IEEE, 2022. — P. 63–68. doi: 10.1109/RusAutoCon54946.2022.9896267
13. Ovchinnikov E.L. Chastotnoe vospriyatie zvuka: Vy'sota tonov v texnike: mel-shkala [Frequency perception of sound: Tone pitch in the technique: chalk scale] / E.L. Ovchinnikov // Science and education in the modern world. — 2016. — 11. — P. 39–42. [in Russian]
14. Huang C.L. Exploring effective data augmentation with TDNN-LSTM neural network embedding for speaker recognition / C.L. Huang. // 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU); — 2019: IEEE, 2019. — P. 291–295.