

# ПРОГРАММНЫЙ КОМПЛЕКС РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД ДЛЯ ДИКТОРОВ С НАРУШЕНИЕМ РЕЧИ

Зарудный Д.И., Лешенко Ю.П., Серов А.А., Симонов Е.В.

*Московский технологический университет (МИРЭА), 119454, Россия, г. Москва, проспект Вернадского, 78, [mirea@mirea.ru](mailto:mirea@mirea.ru)*

---

**Профессиональное образование играет определяющую роль в профессиональной реабилитации инвалидов, т.к. именно оно создает основу для реализации принципа равных возможностей и инклюзивного образования. Существующие русифицированные зарубежные программы речевого ввода ориентированные на английский язык, а используемые методы, не учитывают специфику дикторов с нарушением речи. Представленный программный комплекс распознавания речевых команд позволяет облегчить взаимодействие с компьютером пользователей с нарушениями опорно-двигательной системы, а также, незрячих или слабовидящих.**

---

Ключевые слова: распознавание, программы речевого ввода, программный комплекс распознавания речевых команд.

## PROGRAM COMPLEX OF RECOGNITION OF SPEECH TEAM FOR SPEAKERS WITH VIOLATION OF SPEECH

ZARUDNY D.I., LESHCHENKO YU.P., SEROV A.A., SIMONOV E.V.

*Moscow Technological University (MIREA), 119454, Russia, Moscow, Vernadsky Prospect, 78, [mirea@mirea.ru](mailto:mirea@mirea.ru)*

---

**Vocational education plays a decisive role in the vocational rehabilitation of disabled people. It creates the basis for the realization of the principle of equal opportunities and inclusive education. Existing Russified foreign speech input programs oriented to English, and the methods used, do not take into account the specifics of speakers with speech disorders. The presented subsystem of recognition of speech commands allows to facilitate interaction with a computer of users with disorders of the musculoskeletal system, as well as, blind or visually impaired.**

---

Keywords: recognition, speech input programs, speech recognition subsystem

### 1. Актуальность и постановка задачи

Социальная адаптация и реабилитация лиц с ограниченными возможностями здоровья во всем мировом сообществе является одной из самых актуальных задач. Важнейшим аспектом реабилитации инвалидов является их профессиональная реабилитация. Профессиональное образование играет определяющую роль в профессиональной реабилитации инвалидов, т.к. именно оно создает основу для реализации принципа равных возможностей инвалидов с не инвалидами. Существующие русифицированные зарубежные программы речевого ввода ориентированные на английский язык, а используемые методы, не учитывают специфику дикторов с нарушением речи. Программный комплекс распознавания речевых команд (РПК) позволяет облегчить взаимодействие с компьютером пользователей с нарушениями опорно-двигательной системы, а также, незрячих или слабовидящих. Программный комплекс РПК облегчает доступ лиц с ограниченными возможностями здоровья к информационно-образовательным ресурсам с использованием информационных технологий, способствует реализации психотерапевтической программы, в частности, повышения уровня самооценки, формирование чувства самостоятельности, ответственности, формирование полноценной учебной деятельности.

Для решения поставленной задачи необходимо разработать программный комплекс РПК для дикторов с нарушением речи и программный модуль (агент), реализующий сценарии выполнения речевых команд. Интеграция программного комплекса РПК с ОС – «Агент обработки сценариев» (АОС), позволяет комплексно подойти к решению задачи профессионально-образовательной реабилитации лиц с ограниченными возможностями здоровья. Круг пользователей данной разработки - Центры психолого-медико-социального сопровождения детей и подростков, высшие и средние учебные заведения, занимающиеся профессиональной подготовкой лиц перенесших ДЦП, с нарушениями зрения или дефектами речи.

Таким образом, с учётом изложенного выше, задача разработки программного комплекса РПК для дикторов

с нарушением речи и программного модуля (агента), реализующего сценарии выполнения речевых команд, является актуальной. Для решения поставленной задачи необходимо рассмотреть следующие подзадачи:

- проблемы и методы распознавания речевых команд для дикторов с нарушениями речи;
- структура и функционирование программного комплекса РПК;
- методология управления приложениями компьютеров с использованием речевого интерфейса

## **2. Математическое обеспечение программного комплекса РПК**

Проанализированы методы распознавания речевых команд и даны рекомендации по их использованию, объединению положительных сторон различных подходов к распознаванию речи. Апробированы методы динамического программирования для распознавания фонематических категорий, модели распознаваемых команд описываемые в виде скрытых Марковских моделей.

Проведены исследования на базе созданного прототипа программного комплекса РПК, как речевого материала выбранной группы дикторов, так и возможностей выбранных методов и алгоритмов.

Использован подход распознавания речи, основанный на механизме предварительной классификации звуков, который, в отличие от традиционного подхода, не требует индивидуальной настройки программы для каждого пользователя, что особенно важно для случая удалённых пользователей с дефектами речи. Модуль предобработки речевого сигнала позволяет значительно повысить качество предобработки, позволяет системе работать с меньшей зависимостью от акустических параметров аудитории и применяемых микрофонов. Кроме того, даёт дополнительный набор параметров речевого сигнала, на основе которых возможно повысить качество распознавания системой в целом. Примененный в программном комплексе РПК метод автосегментирования речи на фонемоподобные звуки позволяет работать с речью отличающейся сильной вариативностью длительности произношения, что характерно для дикторов с нарушениями речи.

Использован и исследован новый подход распознавания речи: механизм предварительной классификации звуков, который, в отличие от традиционного подхода, ослабляет требование индивидуальной настройки программы для каждого пользователя, что особенно важно для случая удалённых пользователей с дефектами речи. Это способствует, так же, повышению помехозащищённости при распознавании речи. Для разработки словаря команд и метода для их надежного распознавания проведён экспертный анализ записей речи дикторов с характерными нарушениями. Анализ включает следующие этапы: получение речевого материала разных дикторов, по возможности, с разной степенью нарушений; определение основных групп звуков, таких как шумовые, тоновые, взрывные или какие-то другие группы, которые присутствуют в речи дикторов; определение частотных, амплитудных и временных параметров характерных для каждой из таких групп. Исследована эффективность системы РПК с использованием методов динамического программирования для распознавания фонематических категорий и моделей распознаваемых команд, описываемых в виде скрытых Марковских моделей. Целесообразно объединить положительные стороны двух различных подходов к распознаванию речи.

Для повышения эффективности распознавания используется классификация речевых патологий: зубочелюстные аномалии – косноязычие, дизартрия - детский церебральный паралич, заикание, выявлены характерные признаки и составлен перечень задач реабилитации лиц для каждого класса речевых патологий. Показано, что по всем видам речевых аномалий можно, используя программный комплекс РПК, количественно оценить степень приближения речи диктора к эталону, что делает возможным определение наиболее эффективных методов реабилитации, осуществление мониторинга для оценки динамики поэтапной терапии. На базе программного комплекса РПК разработана методология реабилитации лиц с аномалиями речи.

## **3. Структура и функционирование программного комплекса РПК**

Программный комплекс РПК является, полнофункциональным программным продуктом, предназначенным для исследований специфики речевых команд дикторов с нарушением речи, подбора набора команд для наиболее надежного распознавания, формирования словаря, для создания речевого интерфейса в компьютерных системах, проверки общей концепции построения системы, проведения проектных исследований и для настройки параметров применяемых алгоритмов.

Программный комплекс РПК включает следующие основные модули:

**модуль предобработки включает следующие блоки:**

- оцифровка,
- фреймирование,
- вычисление характеристических векторов.

Используются алгоритмы, позволяющие проводить более качественный анализ речевого сигнала, вычислять

больше параметров используемых при распознавании и позволяющие значительно снизить влияние типа и качества микрофона на результаты распознавания, использовать библиотеки предобработки не только для распознавания речевых команд, но и для накопления и хранения речевой информации в других задачах дистанционного обучения;

**Модуль классификации фреймов**, используемый для выявления речевых команд в потоке аудио данных, а также для выделения вокализованных и невокализованных участков речевого сигнала;

**Модуль кластеризации**, предназначенный для сегментирования речевого сигнала конечной длины на фонемоподобные звуки;

**Модуль хранения эталонов** состоит из нескольких массивов структур. Каждый массив структур содержит эталоны определенного класса. Используется два класса эталонов, один для вокализованных звуков и один для невокализованных;

**Модуль хранения моделей речевых команд**, представляемых в системе в виде дискретных моделей скрытых Марковских процессов. В качестве дискретных символов процесса используются эталоны фонемоподобных звуков представленные в модуле хранения эталонов;

**Модуль обучения** включает следующие блоки:

- процедура выравнивания,
- сопоставление с эталонами,
- создание нового эталона,
- определение параметров состояний;

**Модуль распознавания** включает следующие блоки:

- процедура распознавания, которая сводится к известной forward- процедуре и не представляет собой сложности,
- процедура обработки ошибок распознавания. С помощью этого блока можно повышать качество распознавания системы в процессе работы с ней и организовать коррекцию системы под другого диктора.

Исходные данные Программный комплекс РРК принимает как непосредственно с микрофона, так и из wav файлов, заранее подготовленных. Предоставляется возможность сохранять промежуточные данные в различных форматах, в том числе, речевые единицы, записанные с микрофона, можно сохранить в виде wav файлов для проведения повторных экспериментов. Речевые единицы, с которыми работает Программный комплекс РРК, могут быть сохранены на внешнем носителе не только в виде wav-файла, но и виде rav-файла, который может содержать представление речевой единицы, как в виде последовательности fft-коэффициентов, так и виде последовательности кепстральных коэффициентов. Эта возможность значительно расширяет сферу применения подсистемы РРК, поскольку она может применяться для получения исходных данных для анализа другими программными продуктами, например, такими как: MATLAB, STATISTICA и пр. Программный комплекс РРК сохраняет полученные модели в виде форматированных текстовых файлов, которые затем могут быть использованы в качестве исходных данных для других систем анализа и распознавания речи. Исходный текст работающего прототипа подсистемы РРК написан на языке C++ как мультитредовое приложение с применением среды разработки приложений Borland C++ Builder 4.0. Имеется программная документация эксплуатационная и сопровождения.

Реализация подсистемы РРК в виде нескольких взаимодействующих тредов позволяет наиболее эффективно использовать ресурсы системы и даёт принципиальную возможность осуществить работу со звуковой подсистемой компьютера в режиме реального времени.

Передача данных между тредом осуществляется посредством очередей и синхронизированного доступа к ним. Данные в очередях представляются в виде структур, содержащих всю необходимую и сопутствующую информацию для правильной работы последующих модулей-тредов. Возможно, в результате исследований будет выявлена необходимость пересмотра и переработки некоторых алгоритмов системы распознавания.

Лицензионные прикладные программные средств в данной работе не используются. Кроме своего прямого назначения Программный комплекс РРК обладает следующими возможностями:

- проведение исследований специфики речевых команд дикторов с нарушением речи;
- подбор команд для наиболее надежного их распознавания;
- формирования словаря.

Предоставляется открытый доступ к коду разработанного подсистемы РРК, что позволяет легко ее модернизировать и привлекать к разработке дополнительные людские ресурсы, что значительно снижает риски проекта в целом.

Программный комплекс РРК включает следующие модули:

**Модуль предобработки** включает следующие блоки:

- оцифровка,
- фреймирование,
- вычисление характеристических векторов.

Используются алгоритмы, описанные в стандарте ETSI ES, что позволяет проводить более качественный анализ речевого сигнала, вычислять больше параметров используемых при распознавании и позволяет значительно снизить влияние типа и качества микрофона на результаты распознавания. Это так же позволяет использовать библиотеки предобработки не только для распознавания речи, но и для накопления и хранения речевой информации в других задачах дистанционного обучения;

**Модуль классификации фреймов**, используемый для выявления речевых команд в потоке аудио данных, а также для выделения вокализованных и невокализованных участков речевого сигнала;

**Модуль кластеризации**, предназначенный для сегментирования речевого сигнала конечной длины на фонемоподобные звуки;

**Модуль хранения эталонов** состоит из нескольких массивов структур. Каждый массив структур содержит эталоны определенного класса. В прототипе используется два класса эталонов, один для вокализованных звуков и один для невокализованных. Количество классов можно увеличить для повышения гибкости системы, в частности, можно добавить классы эталонов по разным частотам основного тона (мужские/женские голоса);

**Модуль хранения моделей команд**. Модели команд представляются в системе в виде дискретных моделей скрытых Марковских процессов. В качестве дискретных символов процесса используются эталоны фонемоподобных звуков представленные в модуле хранения эталонов;

**Модуль обучения** включает следующие блоки:

- процедура выравнивания;
- сопоставление с эталонами;
- создание нового эталона;
- определение параметров состояний. Параметрами состояния является матрица вероятностей генерации различных символов (в данном случае эталонов) в заданном состоянии;

**Модуль распознавания** включает следующие блоки:

- процедура распознавания в данном случае сводится к известной forward-процедуре и не представляет особой сложности;
- обработка ошибок распознавания. С помощью этого блока, т.е. при организации обратной связи, можно повышать качество распознавания системы в процессе работы с ней и организовать коррекцию системы под другого диктора;

**Модуль классов данных** включает:

- класс SPFrame предназначен для хранения информации об одном фрейме акустических данных. Параметры экземпляров этого класса заполняются процедурой считывающей данные из файла или принимающей их со звуковой карты компьютера;

- класс SPFrame является входными данными для процедуры таксономии, которая формирует экземпляры класса TTLWords;

- класс TTLWords содержат информацию о целом слове или его части;

- класс TClasteredWord используется для таксономии слов представленных экземплярами TTLWords на фонемоподобные звуки.

- класс DSPthread реализует интерфейсы Learning и Recognition. Процедура этого класса buildModel, реализующая интерфейс learning, в качестве входных данных принимает  $N > 1$  варианта обучаемой команды в виде экземпляров

- класса TTLWords, формирует соответствующие экземпляры TClasteredWord, и формирует на их основе один экземпляр класса модель и необходимое количество экземпляров класса Etalon.

Класс Etalon содержат информацию о псевдофонамах и являются параметрами класса storeEtalons, который реализует абстракцию хранилища эталонов.

Класс Model содержат информацию о распознаваемых командах и являются параметрами класса store Model, который реализует абстракцию хранилища моделей команд.

Класс DSPthread также реализует интерфейс Recognition. На вход процедура распознавания подается экземпляр неизвестного слова в виде экземпляра класса TTLWords. Далее неизвестное слово кластеризуется и сравнивается со всеми моделями, хранящимися в хранилище моделей.

На Рис.1 представлена функциональная структура подсистемы РРК.

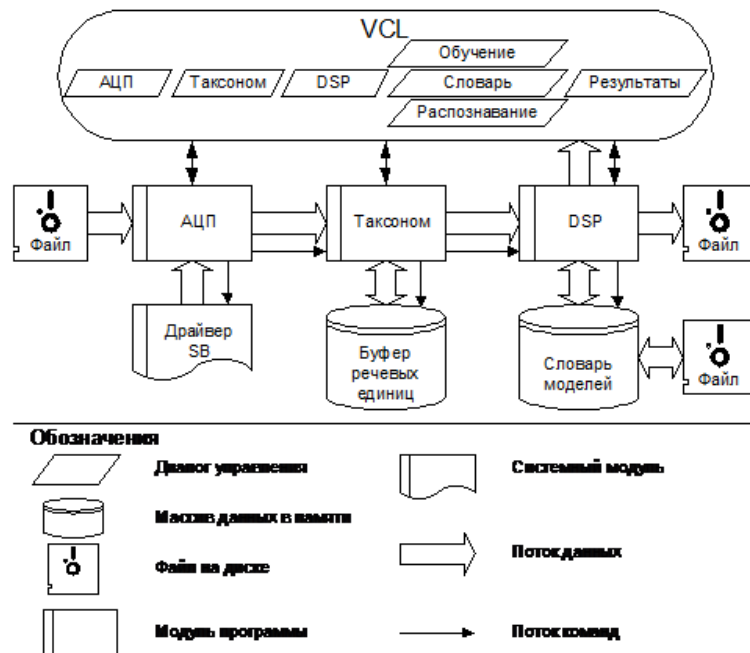


Рис. 1. Функциональная структура программного комплекса

Здесь видны потоки команд и данных между модулями-тредами и их направление. По этой схеме можно судить о маршрутах прохождения информации через модули. Так, например, чтобы данные попали из файла в буфер речевых единиц, они должны проследовать через модуль АЦП, в котором они преобразуются в последовательность фреймов в соответствии с настройками модуля. И наоборот, чтобы данные попали из буфера речевых единиц в файл того или иного формата, они должны проследовать через модуль DSP.

Каждому треду соответствует одно или несколько диалоговых окон, через которые осуществляется управление алгоритмами, которые реализует тред, и так же через эти окна производится отображение результатов работы.

#### 4. Назначение модулей

**VCL.** Этот модуль является основным тредом приложения, правда при этом, он не участвует ни в каких вычислениях. Он занимается взаимодействием с операционной системой и графической подсистемой компьютера в частности.

**АЦП** В обязанности этого модуля-треда входит:

- управление драйвером звуковой карты
- реакция на прерывания со стороны звуковой карты.
- формирование заданного потока фреймов из данных поступающих от звуковой карты.
- отображение параметров текущего сигнала.

В диалоге управления этим тредом можно задать частоту дискретизации и разрядность оцифровки сигнала с микрофона. Здесь же задаются параметры формируемого потока фреймов.

Формируемый поток фреймов снабжается дополнительными параметрами сигнала, необходимыми для таксономического деления речи. К этим параметрам относится энергия сигнала и параметр ZCR (количество пересечений сигналом нуля).

Так же в диалоговом окне этого модуля отображается текущее значение параметров входного сигнала, что позволяет настраивать систему на оптимальную работу с данной звуковой подсистемой компьютера.

#### **Таксоном**

Модуль таксономии речи занимается вычленением изолированных слов из непрерывного потока данных, поступающего от АЦП, и помещает эти слова в буфер речевых единиц. Так же через диалог этого модуля осуществляется управление этим буфером.

По сути, это центр управления основными функциями системы. Отсюда речевые единицы отправляются на предобработку, на обучение и распознавание. Отсюда слова передаются в простейший редактор, встроенный в SDIAPP.

## DSP

Это «главный вычислительный центр». Этот модуль производит все трудоемкие вычисления, те, которые в аппаратных системах обычно выполняет цифровой сигнальный процессор.

Этот модуль обслуживается несколькими диалоговыми окнами, такими как: DSP, Обучение, Распознавание и Словарь.

Поскольку со словарем моделей может работать только модуль DSP, то все обязанности по ведению словаря возложены на этот модуль, и словарь располагается в области данных именно этого треда.

Данные в модуль DSP поступают из модуля Таксоном в виде целых речевых единиц. Далее, в соответствии с «сопроводительной» информацией, модуль производит одну или целую последовательность из ниже перечисленных операций:

- быстрое преобразование Фурье
  - кепстральное преобразование (с возможностью нормализации коэффициентов и вычисления «дельта-расширения»)
  - формирование НММ со стартовыми параметрами, по одному из нескольких алгоритмов
  - вычисление вероятностей для всех имеющихся в словаре моделей, по двум алгоритмам - Forward и Viterbi.
- Сформированные модулем НММ помещаются в словарь моделей, а все остальные результаты либо записываются в файл, либо выводятся в окно Результаты.

Программа подсистемы РПК имеет возможность сохранить все текущие параметры в файл или восстановить их из файла. При работе с программой следует учитывать, что результаты работы программы могут оказаться неверными или даже привести к системным ошибкам, если не учитывать соответствие текущих параметров с теми, при которых проводилось создание моделей. По этому, при создании моделей, целесообразно сохранить файл с настройками системы и затем, при работе с моделями, возвращаться к нему.

Прототип подсистемы РПК был протестирован на тестовом словаре из 45 слов. Качество распознавания прототипа соответствует 63%. Низкий уровень качества распознавания показывает необходимость тонкой настройки и отладки алгоритмов, используемых при распознавании.

Для проведения настройки прототипа системы необходимо:

- провести обширные исследования речевого материала предполагаемой группы дикторов;
- определить параметры длительности произношения фонем;
- определить перечень фонематических категорий позволяющий организовать наиболее надежное их распознавание;
- разработать полный перечень управляющих команд, которые будет использоваться в окончательном варианте специализированного учебного рабочего места (СУРМ);
- для каждой управляющей команды подобрать речевую команду в виде отдельного слова. Таким образом, чтобы в этом слове использовалась максимально отличающаяся от всех остальных речевых команд, последовательность выбранных фонематических категорий;
- на основе накопленного речевого материала создать с помощью подсистемы РПК, файл с эталонами всех выбранных фонематических категорий и файл с моделями выбранных речевых команд;
- провести испытания прототипа с привлечением дикторов из группы, на которую ориентирована разрабатываемая система;
- при необходимости, выявить требования к процессу до обучения системы под каждого конкретного пользователя;
- при необходимости доработать прототип подсистемы РПК, внедрив в него процедуры коррекции под конкретного диктора.

Ключевым комплексом структуры взаимодействия подсистемы РПК с прикладными программами является «Агент обработки сценариев» (АОС).

## 5. Методология управления программным комплексом с использованием речевого интерфейса

Для обеспечения взаимодействия подсистемы РПК с непосредственно управляемым комплексом программных средств используется программный модуль «Агент обработки сценариев» (АОС). Модуль АОС должен быть максимально универсальным, гибким и настраиваемым. Кроме того, модуль АОС используется для озвучивания происходящего, т.е. для генерации речевых сообщений. Модуль АОС реализуется в виде сервиса выполняющего координацию взаимодействия подсистемы РПК с прикладными программами и приложениями ОС по реализации сценариев выполнения речевых команд. В простейшем варианте модуль АОС может реализовать жестко заданный алгоритм взаимодействия. В развитом варианте, используются

конфигурационные xml-файлы с описанными сценариями взаимодействия – это обеспечивает гибкость настройки системы и эффективность использования подсистемы РРК.

Для взаимодействия подсистемы РРК и модуля АОС используется технология socket. Модуль АОС подключает TCP-сервер и принимает через TCP соединение команды от подсистемы РРК. Также это TCP соединение используется для передачи в подсистему РРК контекста распознавания, т.е. перед тем как, по сценарию, должна быть распознана некая команда, модуль АОС сообщает подсистеме РРК перечень возможных команд. Например: если по сценарию в настоящий момент, пользователь должен дать однозначный ответ ДА или НЕТ, то модуль АОС сообщает подсистеме РРК множество {ДА, НЕТ} и ждет от подсистемы РРК результат распознавания, который должен принадлежать этому множеству. Использование «контекстов» позволит значительно повысить качество распознавания. Предусмотрено использование «внеконтекстных» команд, например, таких как ВЫХОД, обработка которых не должна зависеть от контекста. Количество таких «внеконтекстных» команд, нужно минимизировать.

Взаимодействие модуля АОС с прикладными программами и приложениями ОС осуществляется на основе технологии COM. Для реализации механизма непосредственного управления программным комплексом задействован механизм OLE automation. Использование механизма OLE Automation . позволяет организовать максимально эффективное управление всей системой. Также позволяет максимально эффективно использовать возможности стандартного программного обеспечения имеющегося в ОС: Outlook, Word, NetMeeting и т.д

Разработана методология интеграции подсистемы РРК с прикладными программами пользователя и приложениями операционной системы (ОС), обеспечивающая максимальную гибкость настройки взаимодействия с другими прикладными программами (рис. 2).

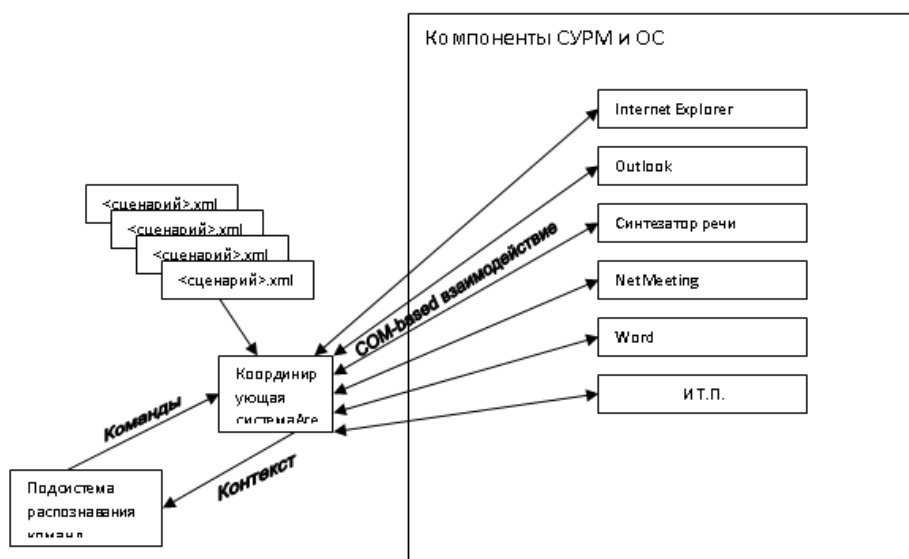


Рис. 2. Схема интеграции программного комплекса с приложениями

Это позволяет наиболее эффективно расширять библиотеку речевых команд и использовать их для управления приложениями:

- реализуется двунаправленный канал передачи данных между системой распознавания и модуля АОС в виде TCP сокета,
- разрабатывается протокол передачи данных по этому сокету,
- отлаживается взаимодействие модуля АОС с синтезатором речи встроенным в операционную систему, т.е. модуль АОС на русском языке озвучивает свои действия;
- управление работой АОС осуществляется с помощью сценариев реализующих команды. В качестве языка написания сценариев, используется XML.

Для реализации механизма управления приложениями используется механизм OLE automation. Этот же механизм применяется и для управления стандартными приложениями.

Алгоритм работы:

- описывается перечень приложений, для которых создаётся речевое управление;
- составляется перечень команд для каждого из этих приложений;

- разрабатывается сценарии управления в виде блок-схем с ветвлениями в местах ввода речевых команд и описанием действий, которые должны быть выполнены на управляемых программных продуктах в ответ на введенные речевые команды;

- разрабатывается и реализуется в АОС программы сценариев
- разрабатываются и реализуются механизмы передачи команд управляемым программным продуктам;
- работа с речью дикторов. Собирается и обрабатывается речевой материал дикторов в соответствии с разработанным перечнем команд в пункте 2. по определённой методике (параллельно пунктам 3-5);
- проводится настройка и отладка системы распознавания речевых команд.

Структура взаимодействия подсистемы РПК с приложениями приведена на Рис. 2

## 5. Заключение

Профессиональное образование играет определяющую роль в профессиональной реабилитации инвалидов, т.к. именно оно создает основу для реализации принципа равных возможностей и инклюзивного образования. Существующие русифицированные зарубежные программы речевого ввода ориентированные на английский язык, а используемые методы, не учитывают специфику дикторов с нарушением речи. Представленный программный комплекс распознавания речевых команд позволяет облегчить взаимодействие с компьютером пользователей с нарушениями опорно-двигательной системы, а также, незрячих или слабовидящих.

## Список литературы

---

1. Липаев В.В. Надежность программных средств. М.: СИНТЕГ, 1998.
2. Зарудный Д.И., Щетинин Д.Ю. Методология автоматизированной разработки информационно-образовательных порталов с учетом стандартов доступности и технологии каркасно-компонентного проектирования. Наука и школа. 2012. № 6. С. 52-55.

## Reference

---

1. Lipaev V.V. Reliability of software. M.: SYNTHEG, 1998.
2. Zarudny D.I., Schetin D.Yu. Methodology for the automated development of information and educational portals, taking into account the standards of accessibility and technology of frame-component design. Science and school. 2012. № 6. P. 52-55.