

Лабораторная работа №2

по теме: Компьютерные информационные технологии  
ввода информации в коммуникационный процесс

# Технология речевого (голосового) ввода информации

Цель работы: ознакомление с компьютерными  
информационными технологиями ввода информации

Работу  
выполнила  
студентка 2  
курса ИАИ ФАД  
Николаева  
Анастасия

# Технология общения с компьютером:

В системах распознавания речи выполняется оцифровка звуковой информации, ее идентификация с кодами, содержащимися в электронных тезаурусных (иногда многоязычных) словарях, необходимая автоматическая коррекция кодов и генерация соответствующих им символов, слов и предложений, возможный вывод текстов на экран для ручной их коррекции (иногда звуковое воспроизведение) и запись текстов в память машины либо исполнение «услышанных» команд.



# Системы которые направлены для распознавания отдельных слов, команд и вопросов.



Системы, ориентированные на распознавание отдельных слов, команд и вопросов часто называют системами речевого управления, поскольку их основная задача - обеспечить выполнение компьютерной системой действий, задаваемых голосом.

Наибольшее распространение такие системы получили в автоматических телефонных службах. В них можно ввести голосом номер телефона вызываемого абонента или его имя; можно задать простой вопрос автоматической справочной службе.

Наиболее разработаны системы распознавания чисел, которые можно отнести к средствам распознавания первого поколения. В развитых системах такого рода человек сначала говорит свой числовой пароль, затем свой числовой идентификатор и только после этого может назвать число, кодирующее сущность запроса.

# Kurzweil Voice for Windows

**Платформа :** Windows 3.1 или выше

**Описание :** Kurzweil Voice for Windows – продукт для диктовки, позволяющий пользователю открывать текст и вводить данные с помощью голоса в windows – приложения. Система адаптивна, но требует настройки.

Пользователь может выбрать активный словарь в 30.000 или 60.000 слов. Трансляция команд поддерживается для множества windows – приложений, таких как: WordPerfect, 1-2-3, Organizer, Word ( список из более чем 30 поддерживаемых приложений приведен на WWW сайте Kurzweil ).

Дополнительная информация содержится там же.

**Системные требования :** 486DX/33 или выше, 8 or 16 МВ памяти ( в зависимости от размера словаря ), 30 МВ дискового пространства, VGA адаптер или выше, саундкарту поддерживаемую Kurzweil.

**Сайт:**

www: <http://www.kurzweil.com/>



# Система распознавания речевых команд "Труфалдино"



## Системы распознавания речи на русском языке

Система является разработкой фирмы "Центр Речевых Технологий "  
(г. Санкт-Петербург)

### **Функциональные характеристики:**

1. распознает разнообразные наборы речевых команд;
  2. активизируется после произнесения ключевого слова;
  3. работает в режиме распознавания для любого языка;
  4. настраивается на любой новый голос;
  5. владеет функцией речевого ответа для подтверждения принятой команды;
  6. имеет отдельную функцию ограничения доступа к управлению - чтобы ограничить доступ пользователей к некоторым командам.
- Эта система ориентирована на распознавание около 100 речевых команд,  
Может быть использована для речевого управления программными комплексами , а также для оперативного ввода цифровой и текстовой информации без использования клавиатуры.

## Системы распознавания речи на русском языке

# Система расознавания команд JUST VOICE 95 PRO (Interactive Products, INC )



Это профессиональная система управления компьютером с помощью голоса для WINDOWS 95 (есть также версия для WINDOWS 3.XX).

Система относится к классу систем распознавания одиночных команд.

Позволяет запускать с помощью голосовых команд клавиатурные и мышинные макросы в любом приложении, при необходимости выполнять команды макросов с регулируемой задержкой. Имеются возможности голосового ответа компьютера в ответ на поданную команду и дополнительного подтверждения пользователем распознанных голосовых команд. Настраивается на говорящего. Любой язык. Устойчивая работа в шумной обстановке. Встроены голосовое подтверждение, голосовой и звуковой ответ. Отслеживается переключение языков и раскладок клавиатуры. Строки из символов национальных алфавитов могут включаться в состав макросов, вызываемых к исполнению голосовыми командами. Непосредственно из макросов могут напрямую вызываться мультимедийные функции. Надёжность распознавания голосовых команд при размере словаря 200 команд - не менее 97%.

# Системы распознавания предложений и связной речи

Системы этой группы делятся на системы отдельной диктовки и системы распознавания связной речи.

Системы отдельной диктовки проще в разработке и технической реализации, но они требуют от пользователя не совсем естественного произнесения фраз — с короткой паузой перед каждым следующим словом. К таким системам относятся, например, ViaType корпорации IBM, Dragon Dictate фирмы Dragon System. Последняя система позволяет, наряду с прочим, непосредственно надиктовывать текст в программы Word, Word Perfect, Internet Explorer, Netscape Navigator и т. д. Активный словарь системы насчитывает десятки тысяч слов и может пополняться пользователем, скажем, по его профессиональной тематике. В системе дополнительно анализируются спектральные (частотные) характеристики каждой буквы, выделяются и хранятся ее отдельные фонемы (элементы спектра). На основе этого анализа создаются фонетические модели букв и формируемых из них слов. Точность распознавания достигает 90 %, а после проверки по словарю еще значительно повышается.

# Системы распознавания предложений и связной речи

Наиболее сложные проблемы возникают при распознавании связной речи. При произнесении связной речи больше сказывается эмоциональная составляющая вводимой информации, и при слитном произношении слов несколько изменяется их звучание — все это, безусловно, затрудняет распознавание. Наиболее продвинутыми системами распознавания слитных текстов можно считать системы распознавания речи: Naturally Speaking Deluxe компании Dragon System, Via Voice корпорации IBM и WildFire фирмы Wildfire Communication, Voice Xpress фирмы Lernoute&Hauspie SpeechProducts. Названные системы позволяют после длительной «тренировки» программы надиктовывать ПК тексты и отдельные команды, иногда даже разными операторами. Так, система ViaVoice позволяет многие виды работ на компьютере выполнять в речевом режиме. Можно надиктовывать текст (письма, отчеты, статьи) непосредственно в Windows-приложения, открывать и закрывать компьютерные файлы, ориентироваться в пределах рабочего стола. Такие речевые команды, как «file save, fale print, scroll up, scroll down» безошибочно выполняются компьютером. Скорость ввода текста достигает 140 слов в минуту, что намного больше средней скорости ввода-информации с клавиатуры.

# Системы идентификации по образцу речи

*Механизм распознавания речи* состоит из четырех основных блоков:

1. препроцессора;
2. экстрактора;
3. компаратора;
4. интерпретатора.

Идентификация по образцу речи относится к биометрическим технологиям идентификации человека по его уникальным физическим признакам, таким как отпечатки пальцев, рисунок радужной оболочки глаз. Речь, подобно подписи, характеризуется множеством постоянных физических параметров (которые, кстати, существенно меньше меняются со временем, чем внешность человека). Цель систем идентификации по образцу речи — идентифицировать конкретного известного системе пользователя и выявить самозванца. Взаимодействие пользователя с системой идентификации состоит из трех этапов: регистрации пользователя с целью запоминания особенностей его голоса и формирования для него речевой модели; тестирования, во время которого выполняется сравнение поступившего образца речи с запомненной речевой моделью пользователя, а также возможное выявление модели самозванца из базы моделей голосов множества прочих людей; допуска к работе в системе, если тестирование прошло успешно, и пользователь назвал верный пароль.



Структурная схема системы автоматического распознавания речи

1. **Препроцессор** или **модуль сбора данных** обеспечивает приведение речевого сигнала к наиболее качественному виду (производится автоматическая регулировка усиления, подавление эхо-сигнала, фиксация наличия или отсутствия речи и интонационного конца фразы).
2. **Экстрактор** выполняет спектральный анализ сигнала. Акустическо-фонетический поток звуков разбивается на короткие кадры (длительностью примерно по 10 мс) и выявляются спектральные характеристики каждого кадра.
3. **Компаратор** выполняет акустическое сравнение выявленных характеристик каждого кадра с имеющимися акустическо-фонетическими образцами. Сравнение производится на уровнях выявления контекстно-независимых фонем, контекстно-зависимых фонем и моделей слов.
4. **Интерпретатор** решает задачу наилучшего разбиения полученного компаратора «алфавитного» потока на слова и фразы.

# Схема распознавания речи компьютером

## Принцип распознавания речи компьютером



**Оцифровка голоса.** На этом этапе качество зависит от чёткости дикции, качества микрофона и звуковой карты.

**Сравнение записи с записями в словаре.** Здесь работает принцип "чем больше – тем лучше": чем больше записанных слов содержит словарь, тем выше шансы того, что Ваши слова будут распознаны правильно.

**Вывод текста.** Система автоматически, ориентируясь по паузам, пытается выделить из потока речи отдельные лексемы, соответствующие шаблонным лексемам из словаря, а затем выводит найденные соответствия в виде текста.

## Выводы:

До сих пор не существует стопроцентного способа, который бы обеспечил полное распознавание голоса в текст. За последние годы наметился заметный прогресс.

Уже сейчас мы используем некоторые наработки, которые распознают речь с точностью от 60 до 90% (зависит от Вашей дикции и качества микрофона). На сегодняшний день такой точности достаточно для надиктовки чернового варианта любого текста с последующей небольшой его правкой.