

**МИНИСТЕРСТВО ПО РАЗВИТИЮ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ И КОММУНИКАЦИЙ РЕСПУБЛИКИ УЗБЕКИСТАН**

**НУКУССКИЙ ФИЛИАЛ ТАШКЕНТСКИЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ИМЕНИ
МУХАММАДА АЛ-ХОРЕЗМИЙ**

КАФЕДРА «ПРОГРАММНЫЙ ИНЖИНИРИНГ»

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС

по дисциплине

АЛГОРИТМЫ РАСПОЗНАВАНИЯ РЕЧИ

| | |
|--------------------------|--|
| Научная сфера: | 300000 – Производственно-техническая сфера |
| Сфера образования: | 330 000 – Компьютерные технологии и информатика |
| Направление образования: | 5330300 – Компьютер инжиниринг |

Нукус -2018

Составила:

Юлдашев К.Р., ассистент кафедры «Программный инжиниринг» НФ ТУИТ имени Мухаммада ал-Харезмий.

Рецензент:

Кожаметов А., доцент кафедры НФ ТУИТ имени Мухаммада ал-Хорезмий.

Рабочая программа утверждена на заседании научно-методического совета факультета Компьютер инжиниринг от «___»_____ 2018., протокол №_____.

Председатель научно-методического совета: _____ д.т.н К.Сеитназаров

СОДЕРЖАНИЕ

1. Учебные материалы.

- 1.1. Темы лекционных занятий.
- 1.2. Темы практических занятий.
- 1.3. Темы самостоятельных работ.
- 1.4. Глоссарий.

2. Приложения.

- 2.1. Учебная программа.
- 2.2. Рабочая программа.
- 2.3. Раздаточные материалы.
- 2.4. Тесты и варианты для контроля знания студентов.
- 2.5. Методические указания по применению критериев оценок.

1.1. Темы лекционных занятий.

| | |
|--|----|
| 1-ЛЕКЦИЯ. Голосовое общение. речь и слух человека..... | 5 |
| 2-ЛЕКЦИЯ. Звуковое оборудование компьютера. преобразование звука в поток чисел | 11 |
| 3-ЛЕКЦИЯ. Анализ речевых сигналов. осциллографические исследования | 17 |
| 4-5-ЛЕКЦИИ. Введение в нейронные сети. структуры нейронных сетей..... | 25 |
| 6-7-ЛЕКЦИИ. Методы распознавания речи. распознавание по образцу | 46 |
| 8-9-ЛЕКЦИИ. Выделение лексических элементов речи..... | 48 |
| 10-11-ЛЕКЦИИ. Предварительная обработка звуковых сигналов | 50 |
| 12-ЛЕКЦИЯ. Выделение информативных признаков речевого сигнала | 52 |
| 13-ЛЕКЦИЯ. Выделение фонем и аллофонов | 53 |
| 14-ЛЕКЦИЯ. Уровни распознавания слитной речи | 56 |
| 15-ЛЕКЦИЯ. Применение нейронных сетей для распознавания речи | 57 |
| 16-ЛЕКЦИЯ. Применение вейвлет-преобразований при распознавании речи ... | 59 |
| 17-ЛЕКЦИЯ. Системы голосового управления | 61 |
| 18-ЛЕКЦИЯ. Методы синтеза речи | 66 |

1-ЛЕКЦИЯ. ГОЛОСОВОЕ ОБЩЕНИЕ. РЕЧЬ И СЛУХ ЧЕЛОВЕКА

План:

1. Голосовое общение .
2. Речи и слух человека .

Уже несколько десятков лет тому назад люди предпринимали попытки создания систем синтеза и распознавания речи. Разумеется, все эти попытки начинались с исследования анатомии и принципов работы речевых, а также слуховых органов человека, в надежде смоделировать их при помощи компьютера и специальных электронных устройств.

В этой главе мы расскажем Вам о том, как примерно устроены органы речи и слуха человека, и как они работают. Конечно, наше изложение анатомических подробностей не претендует на полноту. Оно предназначено лишь для того, чтобы дать Вам некоторое представление о тех вещах, понимание которых имеет значение для систем синтеза и распознавания речи. Более точную и подробную информацию можно найти в любом анатомическом атласе.

Мы также приведем минимальные сведения о такой науке, как языкознание, т.е. науке о языке. Эта наука также называется лингвистикой. В рамках этой книги мы не сможем изложить все постулаты этой науки и ограничимся лишь теми минимальными сведениями, которые пригодятся нам при изучении методов синтеза и распознавания речи.

В этой главе Вы познакомитесь с терминами, которыми оперируют специалисты в области речевых технологий. А главное, сумеете оценить сложность исходных данных, с которыми имеют дело эти специалисты.

Речевой тракт человека

Как и все, созданное природой, речевые органы человека отличаются совершенством. С их помощью человек может не только говорить и петь, но и подражать звукам, издаваемым различными животными.

В этом разделе мы рассмотрим роль в формировании звуков отдельных речевых органов человека, таких как голосовые связки, язык и т.д.

Схема речевого тракта

На рис. 1 мы показали сильно упрощенную схему речевого тракта человека [1]. «Двигателем» этой системы, необходимым для ее функционирования, являются легкие. При выдохе воздух из легких поступает через трахею в гортань, а затем в ротовую и носовую полость.

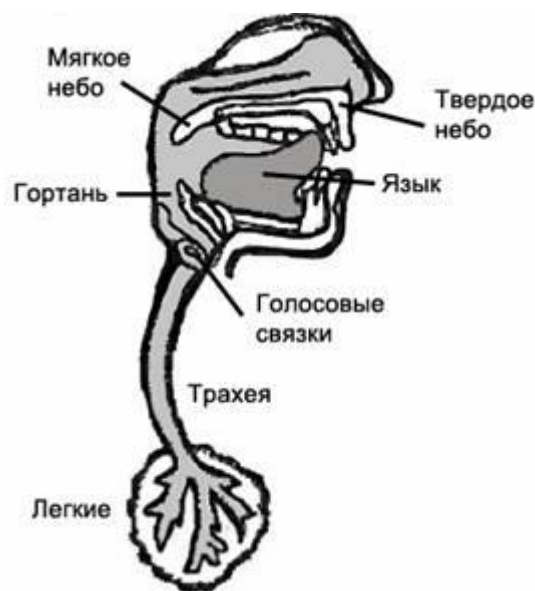


Рис. 1-1. Речевого тракт человека.

Схематически движение воздуха показано на рис. 2. Как видите, на выходе из гортани поток воздуха может раздваиваться, поступая одновременно в носовую и ротовую полость.

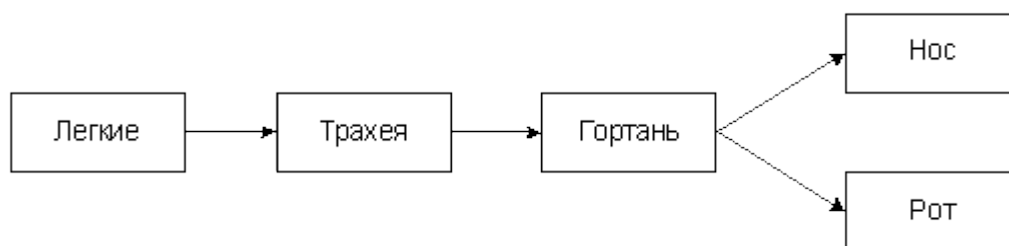


Рис. 1-2. Схема прохождения воздуха при образовании голоса.

Органы, расположенные в ротовой полости, наряду с голосовыми связками, играют решающую роль в формировании звуков. Что же касается носовой полости, то она служит резонатором, усиливая колебания определенных частот.

Голосовые связки

Колебания голосовых связок, расположенных на входе в гортань, наполняет речь человека голосом (мелодичным или не очень, это у кого как). Голосовые связки не работают, когда человек говорит шепотом.

При повреждении голосовых связок в результате болезни или травмы человек способен только шептать. К счастью, специально для этого случая были придуманы электронные устройства с излучателем звука, который прикладывается к горлу и добавляет к шепоту голос. Нельзя сказать, чтобы результаты были потрясающими, так как голос получается тихий и слова бывает трудно разобрать, однако это лучше, чем ничего.

В зависимости от того, какие звуки и как произносит человек, может работать либо одна из полостей, либо обе полости. Носовые звуки произносятся при закрытом рте. Что же касается остальных звуков, то в их формировании принимают участие как носовая, так и ротовая полости. В самом деле, при насморке голос человека заметно изменяется — он начинает говорить «в нос». Из этого следует, что в формировании речи активно участвует и полость рта, и носовая полость.

Активные и пассивные органы речи

Все органы, участвующие в формировании речи, можно разделить на активные и пассивные органы. При этом в процессе речи активные органы совершают различные движения, формируя звуки. Вот активные органы речи:

- ☐ голосовые связки;
- ☐ язык;
- ☐ губы;
- ☐ мягкое небо;
- ☐ язычок;
- ☐ задняя спинка зева;
- ☐ нижняя челюсть.

Пассивные органы речи играют лишь вспомогательную роль. Они, в частности, определяют форму полостей, от которой, в свою очередь, зависят резонансные свойства этих полостей. Следующие органы речи являются пассивными:

- ☐ зубы;
- ☐ альвеолы;
- ☐ твердое небо;
- ☐ верхняя челюсть.

Заметим, что хотя пассивным органам речи отведена вспомогательная роль, их значение нельзя преуменьшать. Отсутствие, например, нескольких зубов нижней челюсти может привести к заметным дефектам речи (шепелявое произношение).

Работа речевого тракта

Итак, мы выяснили, из чего состоит речевой тракт человека. Поступая из легких и проходя через гортань, воздух проходит мимо голосовых связок. Колебания этих связок и создают звук, который мы слышим, когда человек говорит или поет. Многочисленные резонаторы, форму которых человек может изменять при помощи активных органов речи, формируют звуковую окраску голоса.

Теперь мы рассмотрим весь этот процесс подробнее, остановив внимание на некоторых деталях, существенных для систем синтеза и распознавания речи.

Артикуляция

Движения, выполняемые органами речи в процессе произнесения звуков, называются артикуляцией.

Артикуляция является сложным процессом, описание которой охватывает до дюжины различных признаков. Процесс артикуляции состоит из трех фаз [2]:

- ☐ приступ (экскурсия);
- ☐ выдержка;
- ☐ отступ (рекурсия).

Во время приступа артикуляции органы речи переходят из спокойного состояния в положение, необходимое для произнесения данного звука.

Во время фазы выдержки органы речи сохраняют свое положение, необходимое для произнесения текущего звука.

И, наконец, на фазе отступа органы речи переводятся в спокойное состояние.

Голосовые и шумовые звуки

Как мы уже говорили, колебания связок придают голосу звучание. В этом звучании выделяется так называемый основной тон, или тон основной частоты. Значение частоты основного тона зависит от размеров и степени натяжения связок.

У разных людей могут быть разные размеры связок, поэтому тональность голоса разных людей обычно различается. Регулируя натяжение связок в процессе артикуляции, человек может менять частоту основного тона.

Помимо голосовых, человек может издавать и шумовые звуки.

Все шумовые звуки можно разделить на два типа: турбулентные и импульсные [3].

Турбулентные звуки образуются при прохождении звука через сужения речевого тракта. Например, согласные с, ф, х, ц, ч, ш, щ произносятся «без голоса» с использованием турбулентных шумовых звуков.

Импульсные шумовые звуки образуются при резком изменении давления при прерывании струи воздуха. Это происходит, когда Вы произносите такие согласные, как п, к, т, д.

Гармоники

Возможно, Вы знаете, что звук идеально чистого тона содержит колебания только одной частоты. График изменения амплитуды звукового сигнала чистого тона может быть представлен в виде идеальной синусоиды.

На практике, однако, звуки с идеально чистым тоном в обиходе встречаются довольно редко. Если, например, скрипач, пианист и певец возьмут ноту «ля», то

отличия в звучании будет заметно на слух, хотя тон звука во всех трех случаях будет одинаковый.

В чем тут дело?

А дело тут в том, что помимо тона основной частоты, в звуке инструментов и голосе всегда присутствуют так называемые гармоники. Гармоники представляют собой звуки других частот, отличных от основной частоты.

В общем случае любой звук можно представить в виде некоторого бесконечного набора абсолютно чистых звуков различных частот. Совокупность частот таких чистых звуков мы будем называть спектром звука.

Таким образом, практически в любом звуке помимо основной частоты присутствуют и другие частоты спектра, называемые гармоническими составляющими, или просто гармониками. От процентного соотношения гармонических составляющих зависит окраска звука.

Все вышесказанное относится, кстати, не только к звуку, но и к свету. Если Вы думаете, что красный сигнал светофора содержит только один какой-то оттенок красного цвета, то на самом деле это не так. Лампа накаливания излучает целый спектр, а красный фильтр светофора выделяет из этого спектра некоторый диапазон красных оттенков.

Другое дело — лазер. Его излучение является монохромным. Это означает, что в нем присутствует только один цвет. Лазерная указка, которую можно купить в любом киоске, излучает монохромный красный цвет, и в этом она сильно отличается от красного фонаря светофора.

Формантные частоты

Теперь Вы знаете, что даже в чистом, на первый взгляд, звучании ноты, может присутствовать целый спектр звуковых частот.

Голосовые органы человека добавляют к основному тону, формируемому голосовыми связками, дополнительные гармонические составляющие. Эти составляющие придают окраску голоса, по которой Вы, в частности, сможете узнавать речь знакомых Вам людей.

В результате исследований было установлено, что в образовании речи активно участвуют четыре частоты, образующиеся в резонансных полостях речевого тракта [1]. Эти частоты называются формантами.

В процессе артикуляции происходит постоянное изменение амплитуды формантных частот, которое можно обнаружить при помощи программ спектрального анализа. Такие программы позволяют развернуть спектр сигнала во времени, отображая его в трехмерном виде.

На рис 1-3 четко виден формантный состав гласных и и у при произнесении последовательности этих звуков [3]. При переходе от гласной и происходит смещение частоты форманты F2 с 2400 Гц на 784 Гц, а также одновременное ослабление формант F3 и F4.

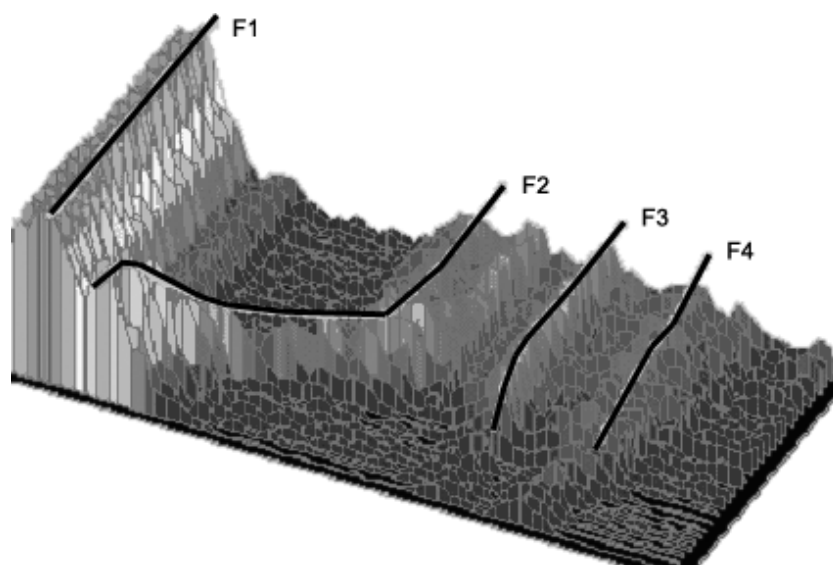


Рис. 1-3. Формантный состав гласных и и у.

Как видите, в процессе артикуляции может изменяться как амплитуда, так и частота формантных составляющих звука. При этом, однако, количество самих формант в голосовых звуках остаётся постоянным и всегда равно 4.

Что же касается шумовых звуков, то в них затруднительно выделить формантные составляющие. Это видно на рис. 1-4, где приведён спектр звука х (представляющего собой турбулентный шум) [3].

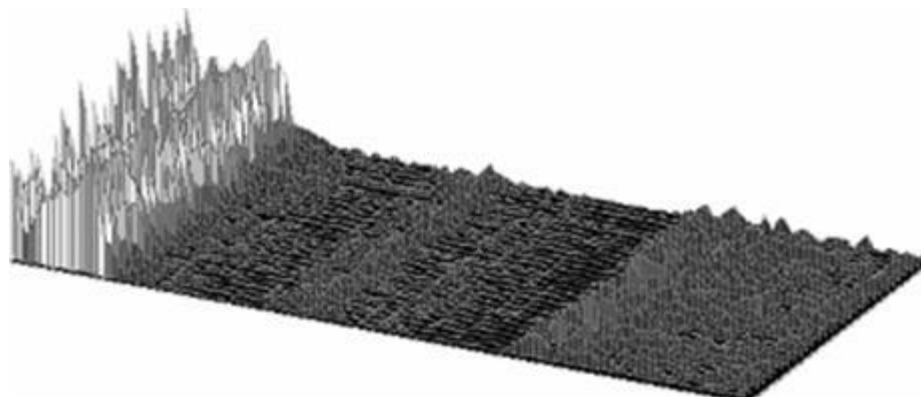


Рис. 1-4. Спектр звука х

Забегая вперед, скажем, что современные системы распознавания речи выполняют спектральный анализ, который позволяет выделить из звуковых сигналов речи наиболее информативные составляющие. Это формантные частоты, а также шум. Помимо спектрального анализа используются и более совершенные методы, такие, например, как вейвлет-преобразования (о них мы расскажем позже в 5 главе).

Что же касается синтеза речи, то один из таких методов предполагает создание звуков речи из отдельных формантных составляющих с добавлением шума.

2-ЛЕКЦИЯ. ЗВУКОВОЕ ОБОРУДОВАНИЕ КОМПЬЮТЕРА. ПРЕОБРАЗОВАНИЕ ЗВУКА В ПОТОК ЧИСЕЛ

План:

1. Звуковое оборудование компьютера
2. Преобразование звука в поток чисел

В этой теме расскажем о том, как ввести в компьютер звуковую информацию, а также как заставить компьютер издавать звуки. Кроме того, мы приведем практические рекомендации по оснащению Вашей лаборатории, в которой Вы сможете самостоятельно провести некоторые исследования речи, а также опробовать существующие системы синтеза и распознавания речи.

Преобразование звука в поток чисел

Во времена, когда исследователи только приступали к решению проблемы создания речевого интерфейса для компьютеров, им зачастую приходилось самостоятельно изготавливать аппаратуру, позволяющую вводить в компьютер звуковую информацию, а также выводить ее из компьютера. Сегодня такие устройства могут иметь разве лишь исторический интерес, так как современные компьютеры можно легко оснастить устройствами ввода и вывода звука, такими как звуковые адаптеры, микрофоны, головные телефоны и звуковые колонки.

Мы не будем углубляться в детали внутреннего устройства этих приспособлений, но расскажем о том, как они работают, и приведем некоторые рекомендации по выбору звуковых компьютерных приспособлений для работы с системами распознавания и синтеза речи.

Микрофон

Как мы уже говорили в предыдущей главе, звук представляет собой ни что иное, как колебания воздуха, частота которых лежит в диапазоне частот, воспринимаемых человеком. У разных людей точные границы диапазона слышимых частот могут изменяться, однако считается, что звуковые колебания лежат в диапазоне 16-20 000 Гц.

Задача микрофона заключается в преобразовании звуковых колебаний в электрические колебания, которые в дальнейшем могут быть усилены, отфильтрованы для удаления помех и оцифрованы для ввода звуковой информации в компьютер.

По принципу действия наиболее распространенные микрофоны делятся на угольные, электродинамические, конденсаторные и электретные. Некоторые из этих микрофонов для своей работы требуют внешнего источника тока (например, угольные и конденсаторные), другие под воздействием звуковых колебаний

способны самостоятельно вырабатывать переменное электрическое напряжение (это электродинамические и электретные микрофоны).

Можно также разделить микрофоны по назначению. Есть студийные микрофоны, которые можно держать в руке или закрепить на подставке, есть радиомикрофоны, которые можно закрепить на одежде, и так далее.

Имеются также микрофоны, предназначенные специально для компьютеров. Такие микрофоны обычно крепятся на подставке, стоящей на поверхности стола. Компьютерные микрофоны могут комбинироваться с головными телефонами, как это показано на рис. 2-1.



Рис. 2-1. Головные телефоны с микрофоном

Как же выбрать из всего многообразия микрофонов тот, что лучше всего подходит для систем распознавания речи?

В принципе, Вы можете экспериментировать с любым имеющимся у Вас микрофоном, если только его можно подключить к звуковому адаптеру компьютера. Однако разработчики систем распознавания речи рекомендуют приобрести такой микрофон, который при работе будет находиться на постоянном расстоянии от рта говорящего.

Если расстояние между микрофоном и ртом не изменяется, то средний уровень электрического сигнала, поступающего от микрофона, также будет меняться не слишком сильно. Это окажет положительное влияние на качество работы современных систем распознавания речи.

В чем тут проблема?

Человек способен успешно распознавать речь, громкость которой меняется в очень широких пределах. Мозг человека способен отфильтровывать тихую речь от помех, таких, например, как шум машин, проезжающих по улице, посторонние разговоры и музыку.

Что же касается современных систем распознавания речи, то их способности в этой области оставляют желать лучшего. Если микрофон стоит на столе, то при повороте головы или изменении положения тела расстояние между ртом и микрофоном будет изменяться. Это приведет к изменению уровня выходного сигнала микрофона, что, в свою очередь, ухудшит надежность распознавания речи.

Поэтому при работе с системами распознавания речи наилучшие результаты будут достигнуты, если использовать микрофон, прикрепленный к головным телефонам, как это показано на рис. 2-1. При использовании такого микрофона расстояние между ртом и микрофоном будет постоянным.

Обращаем также Ваше внимание, что все эксперименты с системами распознавания речи лучше всего проводить, уединившись в тихой комнате. В этом случае влияние помех будет минимально. Разумеется, если Вам нужно выбрать систему распознавания речи, способную работать в условиях сильных помех, то испытания нужно проводить по-другому. Однако, насколько это известно авторам книги, пока помехозащищенность систем распознавания речи еще очень и очень низка.

Аналого-цифровой преобразователь

Микрофон выполняет для нас преобразование звуковых колебаний в колебания электрического тока. Эти колебания можно увидеть на экране осциллографа, однако не спешите в магазин, чтобы приобрести это дорогостоящее устройство. Все осциллографические исследования мы сможем провести с помощью обычного компьютера, оборудованного звуковым адаптером, например, адаптером Sound Blaster. Позже мы расскажем Вам, как это сделать.

Осциллограмма звукового сигнала

На рис. 2-2 мы показали осциллограмму звукового сигнала, получившуюся при произнесении долгого звука а. Эта осциллограмма была получена с помощью программы GoldWave, о которой мы еще расскажем в этой главе книги, а также с помощью звукового адаптера Sound Blaster и микрофона, аналогичного показанному на рис. 2-1.

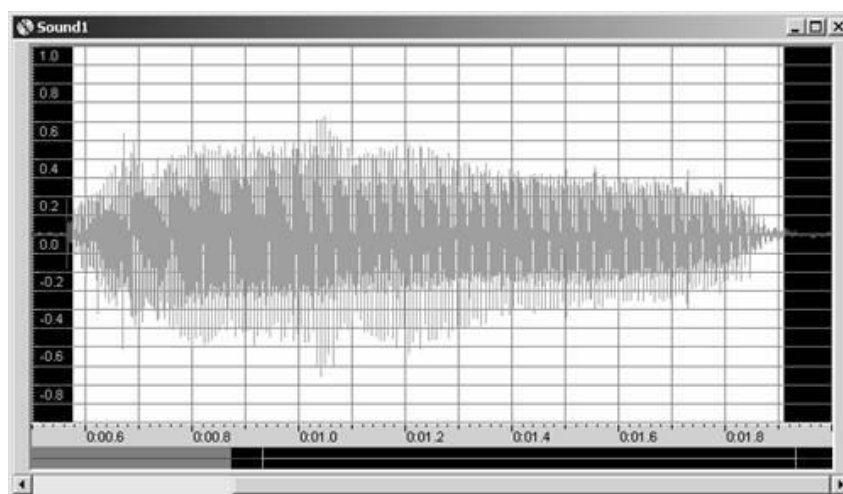


Рис. 2-2. Осциллограмма звукового сигнала

Программа GoldWave позволяет растягивать осциллограмму по оси времени, что позволяет разглядеть мельчайшие детали. На рис. 2-3 мы показали растянутый фрагмент упомянутой выше осциллограммы звука а.

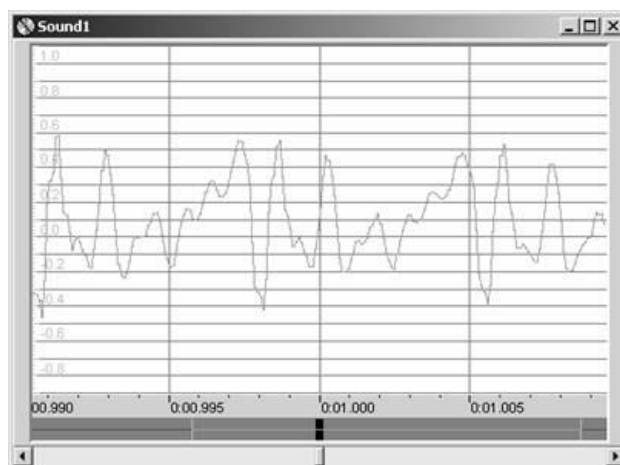


Рис. 2-3. Фрагмент осциллограммы звукового сигнала

Обратите внимание, что величина входного сигнала, поступающего от микрофона, изменяется периодически и принимает как положительные, так и отрицательные значения.

Если бы во входном сигнале присутствовала только одна частота (то есть если бы звук был «чистым»), форма сигнала, полученного от микрофона, была бы синусоидальной. Однако, как мы уже говорили, спектр звуков человеческой речи состоит из набора частот, в результате чего форма осциллограммы речевого сигнала далека от синусоидальной.

Аналоговый и цифровой сигнал

Сигнал, величина которого изменяется со временем непрерывно, мы будем называть аналоговым сигналом. Именно такой сигнал поступает от микрофона. В отличие от аналогового, цифровой сигнал представляет собой набор числовых значений, изменяющихся со временем дискретно.

Чтобы компьютер мог обработать звуковой сигнал, его необходимо перевести из аналоговой формы в цифровую, то есть представить в виде набора числовых значений. Этот процесс называется оцифровкой аналогового сигнала.

Оцифровка звукового (и любого аналогового) сигнала выполняется с помощью специального устройства, называемого аналогово-цифровой преобразователь АЦП (Analog to Digital Converter, ADC). Это устройство находится на плате звукового адаптера и представляет собой обычную с виду микросхему.

Как работает аналогово-цифровой преобразователь?

Он периодически измеряет уровень входного сигнала, и выдает на выходе числовое значение результата измерений. Этот процесс иллюстрируется на рис. 2-4. Здесь прямоугольниками серого цвета отмечены значения входного сигнала, измеренные с некоторым постоянным интервалом времени. Набор таких значений и есть оцифрованное представление входного аналогового сигнала.

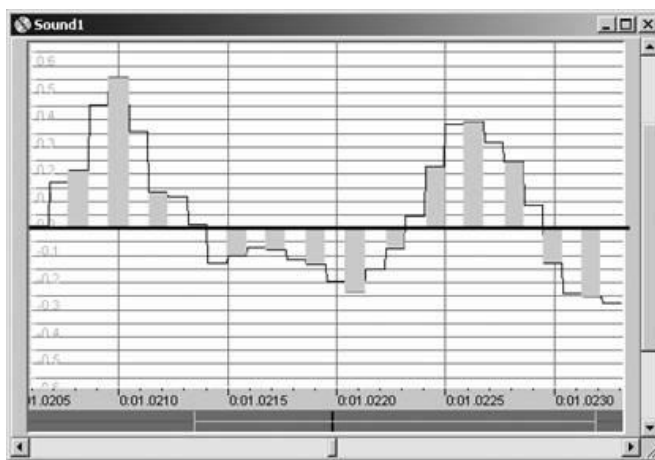


Рис. 2-4. Измерения зависимости амплитуды сигнала от времени

На рис. 2-5 мы показали подключение аналогово-цифрового преобразователя к микрофону. При этом на вход x1 подается аналоговый сигнал, а с выходов u1-un снимается цифровой сигнал.

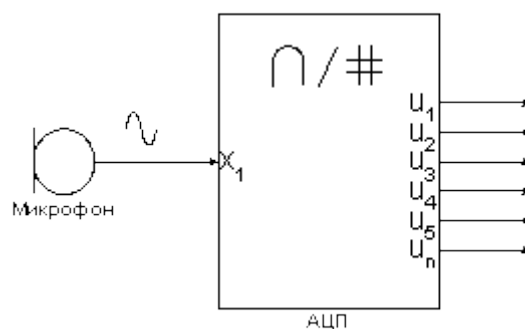


Рис. 2-5. Аналого-цифровой преобразователь

Аналого-цифровые преобразователи характеризуются двумя важными параметрами — частотой преобразования и количеством уровней квантования входного сигнала. Правильный выбор этих параметров критически важен для достижения адекватного представления в цифровом виде аналогового сигнала.

Выбор частоты преобразования

Насколько часто нужно измерять значение амплитуды входного аналогового сигнала для того, чтобы в результате оцифровки не была потеряна информация об изменениях входного аналогового сигнала?

Казалось бы, ответ прост — входной сигнал нужно измерять как можно чаще. Действительно, чем чаще аналого-цифровой преобразователь проводит такие измерения, тем лучше будут отслеживаться малейшие изменения амплитуды входного аналогового сигнала.

Однако излишне частые измерения могут привести к неоправданному росту потока цифровых данных и бесполезной трате ресурсов компьютера при обработке сигнала.

К счастью, правильный выбор частоты преобразования (частоты дискретизации) сделать достаточно просто. Для этого достаточно обратиться к теореме Котельникова, известной специалистам в области цифровой обработки сигналов. Теорема гласит, что частота преобразования должна быть в два раза выше максимальной частоты спектра преобразуемого сигнала. Следовательно, для оцифровки без потери качества звукового сигнала, частота которого лежит в диапазоне 16-20 000 Гц, нужно выбрать частоту преобразования, не меньшую, чем 40 000 Гц.

Заметим, однако, что в профессиональной звуковой аппаратуре частота преобразования выбирается в несколько раз большей указанного значения. Это делается для достижения очень высокого качества оцифрованного звука. Для систем распознавания речи такое качество не актуально, поэтому мы не будем заострять на таком выборе Ваше внимание.

А какая частота преобразования нужна для оцифровки звука человеческой речи?

Так как звуки человеческой речи лежат в диапазоне частот 300-4000 Гц, то минимально необходимая частота преобразования составляет 8000 Гц. Однако многие компьютерные программы распознавания речи используют стандартную для обычных звуковых адаптеров частоту преобразования 44 000 Гц. С одной стороны, такая частота преобразования не приводит к чрезмерному увеличению потока цифровых данных, а другой — обеспечивает оцифровку речи с достаточным качеством.

Выбор количества уровней квантования

Еще в школе нас учили, что при любых измерениях возникают погрешности, от которых невозможно избавиться полностью. Такие погрешности возникают из-за ограниченной разрешающей способности измерительных приборов, а также из-за того, что сам процесс измерений может внести некоторые изменения в измеряемую величину.

Аналого-цифровой преобразователь представляет входной аналоговый сигнал в виде потока чисел ограниченной разрядности. Обычные звуковые адаптеры содержат 16-разрядные блоки АЦП, способные представлять амплитуду входного сигнала в виде $2^{16}=65536$ различных значений. Устройства АЦП в звуковой аппаратуре высокого класса могут быть 20-разрядными, обеспечивая большую точность представления амплитуды звукового сигнала.

Современные системы и программы распознавания речи создавались для обычных компьютеров, оборудованных обычными же звуковыми адаптерами. Поэтому для проведения экспериментов с распознаванием речи Вам не потребуется приобретать профессиональный звуковой адаптер. Такой адаптер, как Sound Blaster, вполне пригоден для оцифровки речи с целью ее дальнейшего распознавания.

3-ЛЕКЦИЯ. АНАЛИЗ РЕЧЕВЫХ СИГНАЛОВ. ОСЦИЛЛОГРАФИЧЕСКИЕ ИССЛЕДОВАНИЯ

План:

1. Анализ речевых сигналов
2. Осциллографические исследования

После установки звукового адаптера и драйвера для него, а также после регулировки чувствительности микрофона, описанной в предыдущей главе, можно провести собственные небольшие исследования в области распознавания и синтеза речи.

В этой главе мы расскажем о некоторых таких исследованиях, которые Вы сможете провести самостоятельно у себя дома при помощи обычного компьютера и доступного программного обеспечения. Разумеется, эти исследования не могут претендовать на полноту и глубину, однако они дадут Вам

некоторое представление о проблеме распознавания речи и возможных путях ее решения.

Осциллографические исследования

Те из Вас, кто когда-либо учился в техническом вузе, знакомы с таким прибором, как осциллограф. Основное назначение этого прибора — визуальное определение формы электрических сигналов и измерение его параметров, таких как амплитуда, частота, длительность импульсов и период их следования и так далее.

Если Ваш компьютер оборудован звуковым адаптером, то при установке специального программного обеспечения Вы сможете использовать его как цифровой осциллограф для изучения электрических колебаний, полученных от микрофона. Таким способом Вы увидите форму звуковых колебаний, образующихся при формировании речи.

Произнося в микрофон отдельные звуки и слова, Вы сможете попытаться разглядеть на осциллограмме отдельные элементы, составляющие речь, проследить изменения амплитуды речевого сигнала и длительности пауз. И хотя информация, полученная в ходе осциллографических исследований, сама по себе не позволит Вам выделить из речи отдельные фонемы и аллофоны, она имеет большое значение для изучения процесса образования речи.

Редактор GoldWave

Для проведения исследований нам, прежде всего, потребуется программное обеспечение, позволяющее просматривать осциллограммы звуковых сигналов. В качестве такого программного обеспечения мы рекомендуем редактор GoldWave (рис. 3-1). Пробную версию этого редактора, работоспособную в течении месяца, можно бесплатно загрузить через Интернет с Web-сайта разработчика <http://www.goldwave.com>.

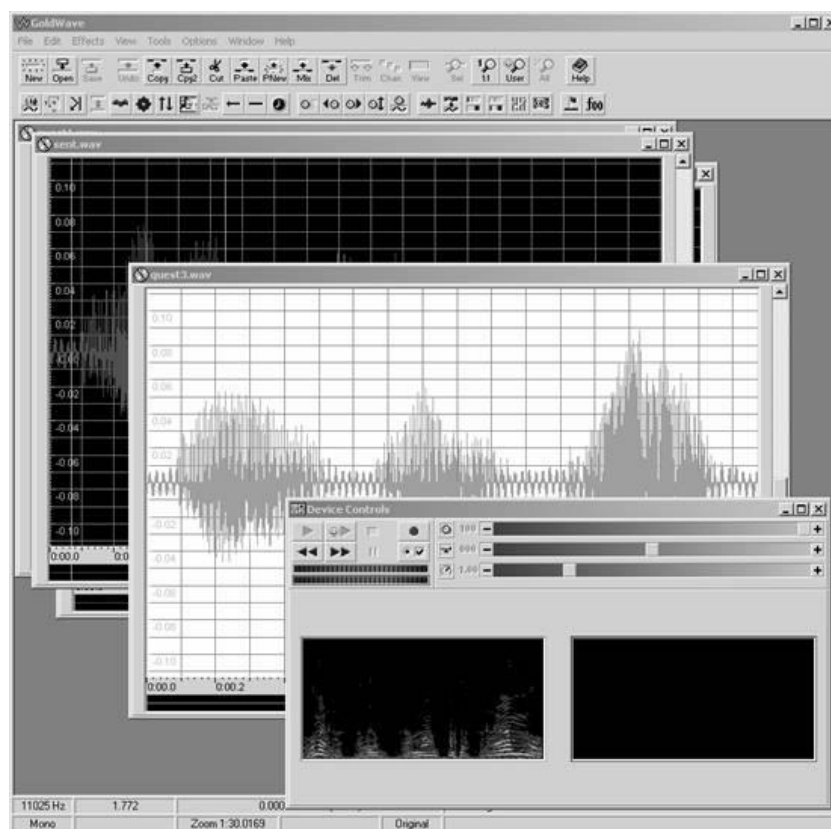


Рис. 3-1. Редактор GoldWave

Программа GoldWave может записывать, проигрывать, редактировать и конвертировать звуковые файлы. Эта программа предназначена в первую очередь для создания высококачественных звуковых файлов, но ее с успехом можно использовать и для исследования речи.

В частности, программа снабжена фильтрами (верхних и нижних частот, а также полосовыми частотными фильтрами). В ней имеется анализатор спектра, который можно использовать для исследования речи.

Пользуясь редактором, можно комбинировать отдельные звуки речи, формируя таким способом синтезированную речь (о том, как устроены синтезаторы речи, мы поговорим позже в 7 главе).

Ниже мы рассмотрим основные приемы работы с программой GoldWave. Подробное описание этой программы можно найти в электронной справочной документации, доступной через меню Help.

Запись звука

В предыдущей главе мы рассказывали о записи звуковых файлов с помощью программы Sound Recorder, входящей в комплект операционной системы Microsoft Windows. Надо сказать, что по сравнению с программой GoldWave возможности Sound Recorder весьма и весьма ограничены.

Для того чтобы записать звук в программе GoldWave, запустите эту программу, а затем выберите из меню File строку New.

На экране появится диалоговое окно New Sound, с помощью которого можно выбрать параметры записи (рис. 3-2).

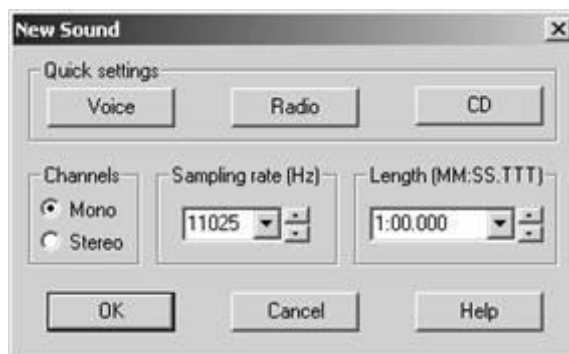


Рис. 3-2. Выбор параметров записи

В группе элементов управления Quick settings предусмотрены кнопки Voice, Radio и CD, при помощи которых можно быстро выбрать параметры звука, соответственно, для записи голоса, радиопередачи или компакт-диска.

Нас интересует запись голоса, поэтому щелкните кнопку Voice. При этом будет создана монофоническая запись (отмечен флажок Mono в группе элементов управления Channels). Также автоматически в списке Sampling rate будет выбрана частота дискретизации входного сигнала, равная 11 025 Гц. Вы можете оставить эту частоту дискретизации, так как она вполне достаточна для оцифровки голоса, либо выбрать при помощи списка Samplingrate другую частоту в диапазоне от 5 500 Гц до 96 000 Гц. Однако не увлекайтесь высокими частотами дискретизации — при обработке речи их применение не даст никакого выигрыша, а приведет лишь к непроизводительным тратам ресурсов компьютера.

При помощи элемента управления Length можно задать длительность записи. Для наших экспериментов достаточно, если запись будет продолжаться одну минуту.

Выбрав параметры записи, щелкните кнопку OK.

Сразу после этого в главном окне программы GoldWave появится новое окно SoundXX, где XX — последовательный номер записей с момента запуска программы, а также окно Device Controls.

Первое из этих окон представляет собой окно звукового редактора, где будет показана осциллограмма звукового сигнала. Во втором окне находятся элементы управления для запуска записи и проигрывания звука, а также дополнительные окна просмотра осциллограммы и спектрограммы записанного сигнала.

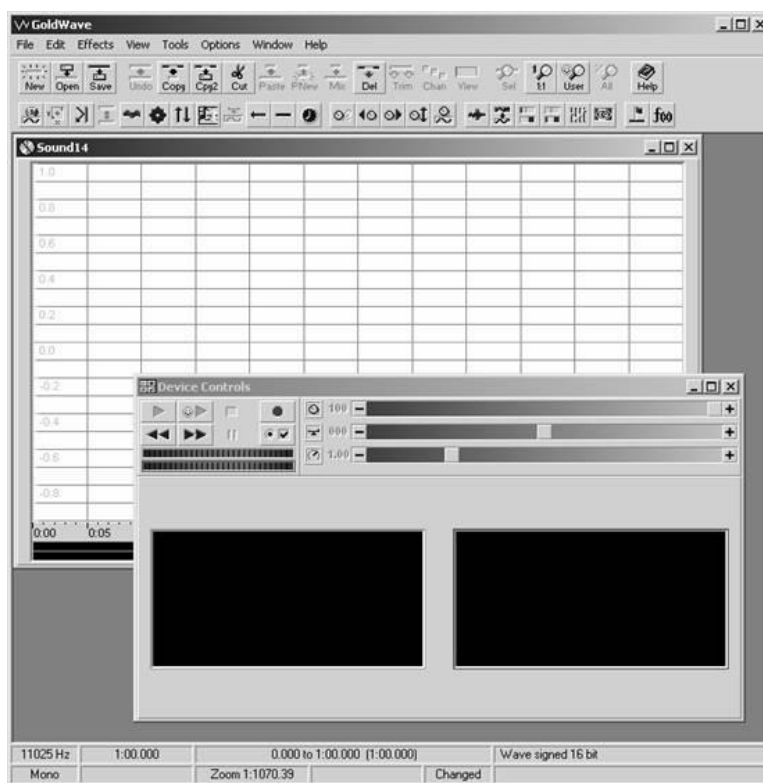


Рис. 3-3. Запись нового звукового файла

Чтобы начать запись, щелкните в окне Device Controls кнопку Record с красным кружком, держа при этом нажатой клавишу Control. Теперь запись включена, и Вы можете произносить слова в микрофон!

Остановка записи

Чтоб остановить запись звука, щелкните в окне Device Controls кнопку Stop. На этой кнопке нарисован квадрат малинового цвета.

Для временной приостановки записи воспользуйтесь кнопкой Pause с изображением двух вертикальных черточек черного цвета.

Сохранение результатов записи в файле

При изучении речи мы советуем Вам записывать слова и фразы в отдельные wav-файлы небольшого размера. В дальнейшем эти файлы можно будет исследовать программой GoldWave или другими программными средствами.

Для сохранения сделанной записи выберите из меню File строку Save As.

В результате на экране появится диалоговое окно Save As, показанное на рис. 3-4. Обратите внимание на список File Attributes, при помощи которого можно выбрать формат создаваемого файла. Для записи речи можно использовать формат, выбранный по умолчанию.

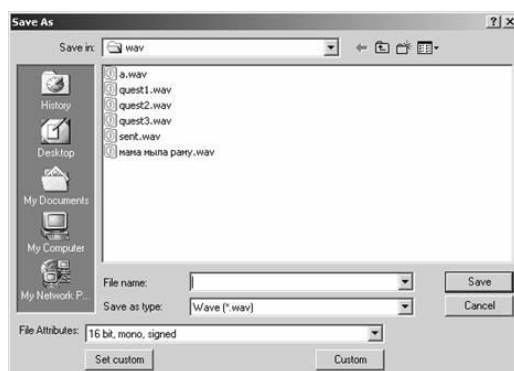


Рис. 3-4. Сохранение звукового файла

Масштабирование

Осциллограмму звукового сигнала, редактируемого в окне программы GoldWave, можно растягивать и сжимать как по вертикали, так и по горизонтали.

На рис. 3-5 мы показали исходную осциллограмму, полученную при записи одного повествовательного и трех вопросительных предложений. Вот эти предложения:

- ☐ Мама мыла раму.
- ☐ `[Мама] мыла раму?
- ☐ Мама `[мыла] раму?
- ☐ Мама мыла `[раму]?

Квадратными скобками отмечены слова, выделенные интонацией при произнесении предложений.

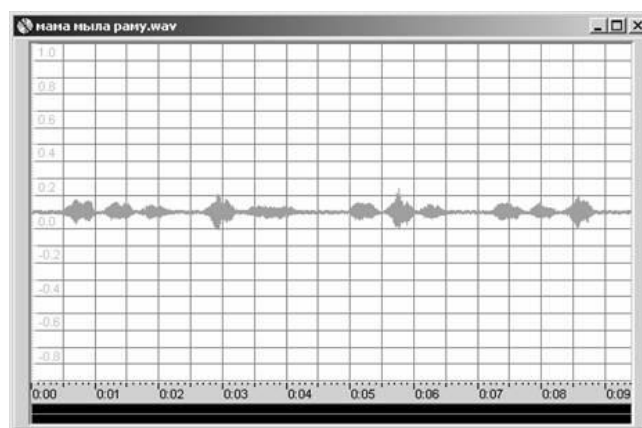


Рис. 3-5. Исходная осциллограмма

Чтобы лучше разглядеть эту осциллограмму, Вы можете растянуть ее по вертикали. Для этого нужно воспользоваться комбинацией клавиши Control и клавиши перемещения курсора вверх по вертикали. Нажимайте эту комбинацию клавиш несколько раз до тех пор, пока не получите нужный масштаб отображения.

На рис. 3-6 мы показали результат растяжения исходной осциллограммы по вертикали. Как видите, теперь на ней появились незаметные ранее детали.

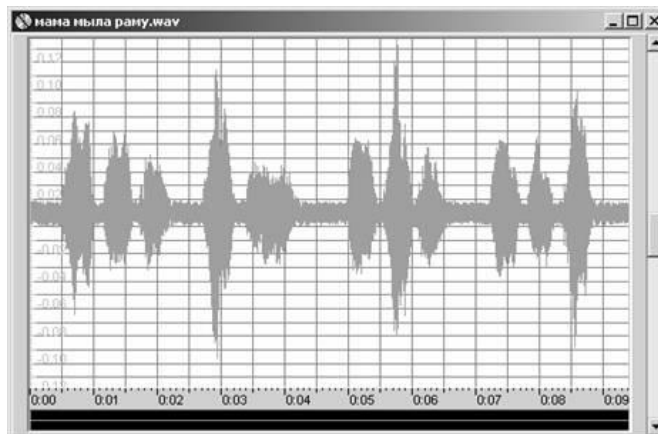


Рис. 3-6. Растяжение по вертикали

Сжатие осциллограммы по вертикали выполняется при помощи комбинации клавиш Control и клавиши перемещения курсора вниз по вертикали.

Чтобы растянуть осциллограмму по горизонтали, воспользуйтесь комбинацией клавиши Shift и клавиши перемещения курсора вверх по вертикали. Результат такого растяжения исходной осциллограммы показан на рис. 3-7.

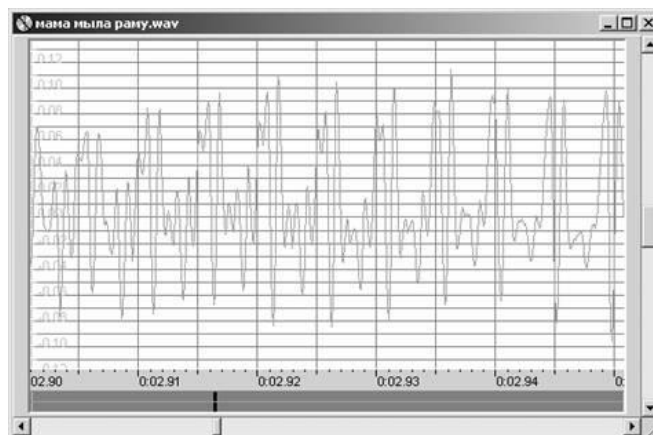


Рис. 3-7. Растяжение по горизонтали

Сжатие осциллограммы по горизонтали выполняется при помощи комбинации клавиш Shift и клавиши перемещения курсора вниз по вертикали.

Нормализация звуковых файлов

Перед проведением дальнейших исследований записанных ранее файлов мы рекомендуем провести нормализацию этих файлов. Нормализация заключается в таком масштабировании амплитуды файла, при котором максимальная амплитуда сигнала будет приведена к максимально возможному значению амплитуды, определяемого разрядностью формата звукового файла.

Чтобы выполнить нормализацию, загрузите в программу GoldWave исходный звуковой файл. Затем выберите из меню **Effects** строку **Volume**, а затем из меню второго уровня — строку **Maximize**.

Выделение фрагмента звукового файла

Одной из часто выполняемых при редактировании звуковых файлов операций является выделение фрагментов с целью их вырезки или перемещения в новый файл.

Выделить нужный фрагмент в окне редактора GoldWave можно с помощью мыши. Для этого вначале нужно щелкнуть левой клавишей мыши в начале выделяемого фрагмента, а затем — правой клавишей мыши в конце выделяемого фрагмента.

Выделенный фрагмент выделяется синим цветом. Это показано на рис. 3-8.

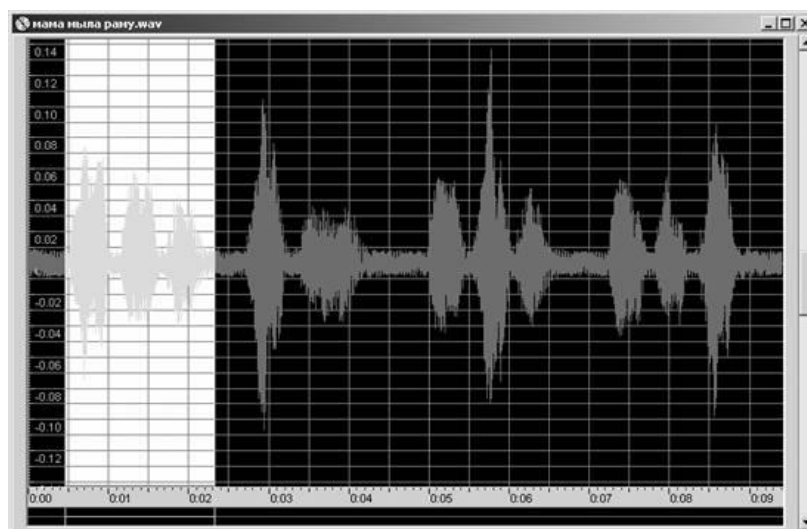


Рис. 3-8. Выделение фрагмента звукового файла

После того как фрагмент выделен, с ним можно выполнять различные операции, доступные через меню **Edit** и **Effects**. Вы можете скопировать фрагмент в универсальный буфер обмена Clipboard, удалить его, наложить фильтр и т.п.

Вырезка фрагмента звукового файла

При проведении собственных исследований речи Вам придется вырезать фрагменты записанных звуковых файлов, сохраняя их в отдельных файлах. Таким способом можно вручную выделять слова, слоги, фонемы, морфемы и другие лексические конструкции.

Операция вырезки фрагмента выполняется очень просто.

Вначале Вам нужно выделить фрагмент файла, используя технику, описанную в предыдущем разделе. Далее надо скопировать выделенный фрагмент в Clipboard, выбрав из меню **Edit** строку **Copy**. И, наконец, нужно создать из содержимого буфера обмена Clipboard новый звуковой файл, выбрав из меню **Edit** строку **Paste new**.

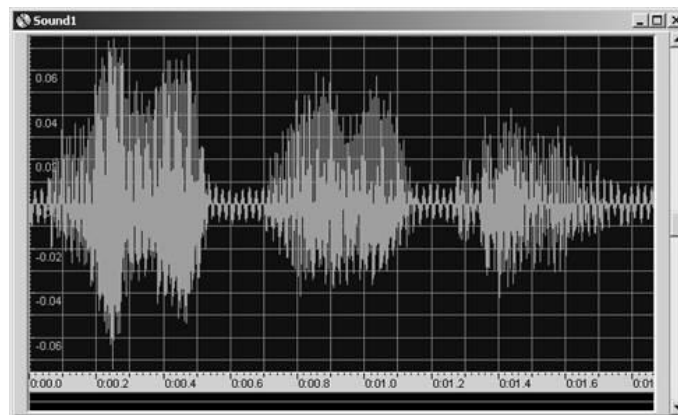


Рис. 3-9. Вырезанный фрагмент звукового файла

Теперь остается только сохранить новый файл, воспользовавшись для этого строкой **Save As** меню **File**.

4-5-ЛЕКЦИИ. ВВЕДЕНИЕ В НЕЙРОННЫЕ СЕТИ. СТРУКТУРЫ НЕЙРОННЫХ СЕТЕЙ

План:

1. Введение в нейронные сети.
2. Структуры нейронных сетей

Одна из важнейших и труднейших задач, решаемых создателями систем распознавания речи, заключается в поиске способа выделения из слитной речи отдельных лексических элементов — фонем, аллофонов, слогов, морфем и т.д. От того, насколько эффективно будет решена эта задача, самым непосредственным образом зависит качество распознавания речи.

К сожалению, попытки понять, каким именно образом справляется с этой задачей человек, пока не привели к особым успехам. Как было сказано в художественном фильме «Формула любви», голова — предмет темный, и исследованиям не подлежит...

Тем не менее, такие исследования, разумеется, идут полным ходом, и они уже дают практические результаты. Здесь мы имеем в виду моделирование нейронов и нейронных сетей, из которых «сделан» головной мозг человека и различных животных.

Искусственные нейронные сети, созданные в виде компьютерных моделей, с успехом справляются с задачами распознавания образов. Они обучаемы, их можно легко приспособить (и это уже сделано) к решению многих практических задач, связанных с распознаванием речи, управлением различного рода машинами и устройствами, прогнозированием событий и пр.

В нашей книге мы не сможем уделить нейронным сетям достаточно внимания и ограничимся лишь описанием основных принципов их действия и обучения. Этот материал необходим, в частности, для чтения следующей главы. В ней мы рассмотрим методы распознавания речи, реализованные в системе анализа речи SAS, о которой мы рассказывали в предыдущей главе.

Тех же, кто желает изучить вопросы проектирования и использования искусственных нейронных сетей более детально, мы адресуем к дополнительной литературе [11]-[15].

Как устроен головной мозг человека

Читая с удивлением статью об очередных успехах полупроводниковой индустрии, сумевшей вместить миллионы транзисторов в одну микросхему центрального процессора, Вы, возможно, и не догадываетесь, насколько человеческий мозг сложнее любого самого современного процессора или суперкомпьютера.

Биологические нейроны

Согласно современным данным, человеческий мозг содержит 1011 живых «компьютеров», называемых *нейронами*. Нейроны соединены между собой в гигантскую сеть, причем общее количество связей между нейронами достигает 1015 [11]. Невозможно даже представить себе, чтобы кто-либо сумел зарисовать схему такого «компьютера», как человеческий мозг, и, тем более, до конца разобрался в том, как эта схема работает.

Тем не менее, удалось более или менее детально изучить устройство и принцип работы элементарной «вычислительной» ячейки мозга — нейрона. Были предприняты многочисленные попытки моделирования биологических нейронов, как с помощью аппаратного обеспечения, так и программными средствами. Нельзя сказать, что на сегодня все тайны биологических нейронов раскрыты, однако созданные модели нейронов неплохо помогают решать многие практические задачи. К таким задачам, в частности, относится распознавание и синтез речи, а также распознавание текста.

Дендриты, синапсы и аксоны

Биологические нейроны представляют собой нервные клетки, соединяемые между собой *дендритами*. Протяжённость дендритов может достигать 1 метра. Точки соединения дендритов называются *синапсами*.

По дендритам сигналы передаются клетке нейрона и принимаются синапсами. Клетка обрабатывает сигналы, поступающие от синапсов, и посылают результаты этой обработки в виде возбуждающего сигнала другим нейронам через *аксон*.

Схема соединения двух нейронов [11] показана на рис. 4-1.

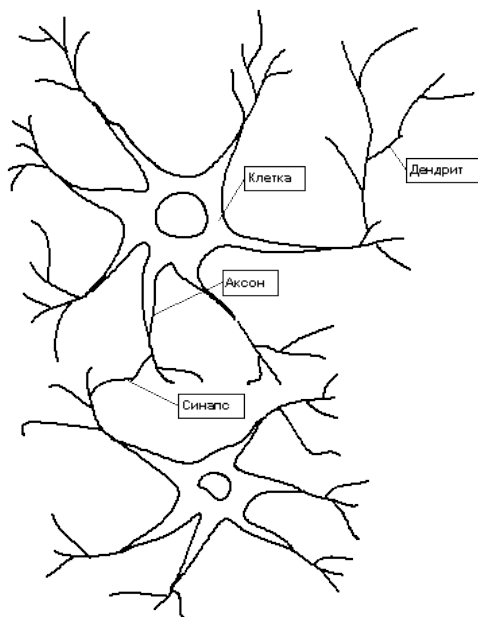


Рис. 4-1. Нейрон, аксоны, синапсы и дендриты

Таким образом, биологический нейрон можно представить себе как устройство с несколькими входами и одним выходом.

Передача возбуждения и торможения

Сигналы, поступающие на входы нейрона (синапсы) складываются с разными знаками. Некоторые из этих сигналов, про которые можно условно сказать, что они положительные, вызывают возбуждение клетки нейрона, другие (условно отрицательные) — торможение.

Когда общее возбуждение клетки нейрона в течение небольшого промежутка времени превышает некоторый порог, клетка вырабатывает сигнал, который передается через ее выход (аксон) другим клеткам.

Надо сказать, что сигналы между нейронами передаются не в виде электрических импульсов, а в виде специальных химических веществ — *нейротрансмиттеров*. Всего насчитывается три десятка различных нейротрансмиттеров. В зависимости от вида вещества может передаваться либо возбуждающий, либо тормозящий сигнал. Что же касается самой передачи сигнала, то она происходит за счет механизма диффузии.

В то же время внутри нейрона наблюдается электрическая активность. Таким образом, работа самого нейрона основана на передаче электрических импульсов.

За более подробной информацией относительно внутреннего устройства биологического нейрона и принципов его работы Вы можете обратиться к [11]. Мы же перейдем к изучению искусственного нейрона, моделирующего (с некоторой степенью точности) работу биологического нейрона.

Искусственный нейрон

На рис. 4-2 мы показали электрическую схему искусственного нейрона, собранную из таких устройств, как умножители, сумматор и функциональный преобразователь.

Входные сигналы поступают в наш искусственный нейрон (в дальнейшем — просто нейрон) через контакты, обозначенные на этой схеме $X_1 \dots X_n$.

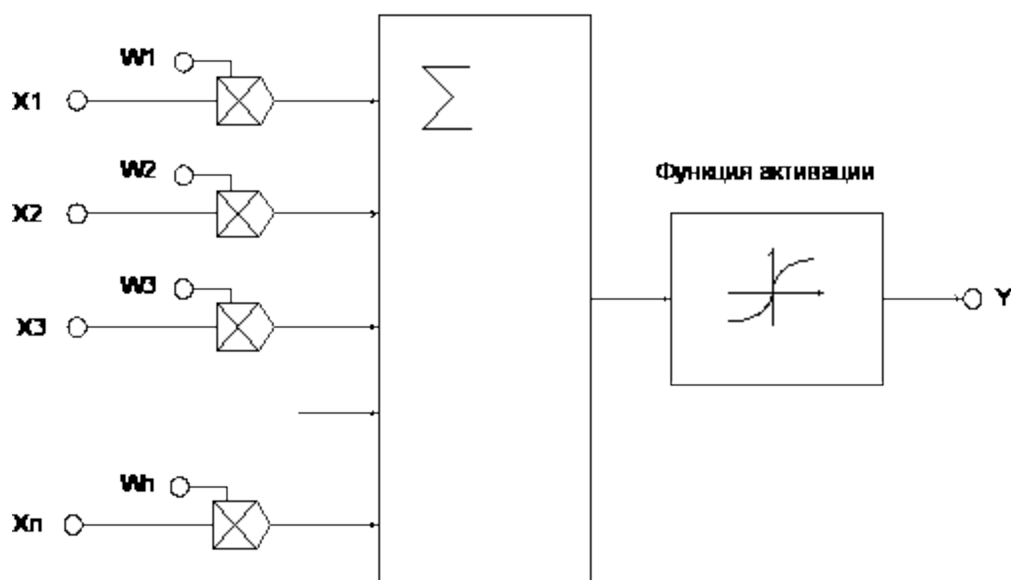


Рис. 4-2. Искусственный нейрон

Входные сигналы умножаются при помощи набора умножителей на некоторые весовые коэффициенты, значения которых подаются на входы $W_1 \dots W_n$.

Далее результаты умножения суммируются, и попадают на вход функционального преобразователя, реализующего так называемую *функцию активации* нейрона.

Рассмотрим эту схему подробнее.

Входы нейрона

В нашей модели входы $X_1 \dots X_n$ играют роль синапсов биологического нейрона. Если модель нейрона создается как электрическая схема, то на эти входы подаются входные сигналы.

Программная модель предполагает запись в «синапсы» нашего нейрона входных значений. При этом входной сигнал нормирован по величине на единицу и может быть как положительным, так и отрицательным числом с плавающей точкой, изменяясь в диапазоне от -1 до 1.

Весовые коэффициенты и память нейрона

Значения входных сигналов умножаются на весовые коэффициенты $W_1 \dots W_n$. Эти коэффициенты позволяют регулировать возбуждение, создаваемое со стороны отдельных входов нейрона.

Таким образом, с помощью весовых коэффициентов мы моделируем различие чувствительности разных синапсов биологического нейрона.

Хотя на рис. 4-2 это не показано, модель нейрона предполагает хранение значений весовых коэффициентов и их изменение. Настраивая весовые коэффициенты нейрона, мы можем изменять его реакцию на тот или иной набор воздействий со стороны входов $X_1 \dots X_n$.

Как Вы увидите дальше, из искусственных нейронов создаются нейронные сети. Эти сети можно обучать нужной реакции на различные возбуждения. Характер этой реакции запоминается именно в виде весовых коэффициентов $W_1 \dots W_n$ отдельных нейронов сети.

В случае программной реализации нейронов память весовых коэффициентов может представлять собой обычный массив, ячейки которого содержат значения коэффициентов.

Сумматор

Сумматор в схеме искусственного нейрона (рис. 4-2) выполняет простую задачу. Он складывает значения, образующиеся при умножении вектора входных значений $[X_1, X_2, \dots, X_n]$ на вектор весовых коэффициентов $[W_1, W_2, \dots, W_n]$.

Величину сигнала на выходе сумматора можно рассчитать по следующей несложной формуле:

$$X_1 * W_1 + X_2 * W_2 + \dots + X_n * W_n$$

Сигнал на выходе сумматора учитывает значения всех входных сигналов, умноженные на соответствующие весовые коэффициенты.

Функция активации

Сигнал, полученный от сумматора, передается функции активации. Результат, полученный при вычислении этой функции, является выходным сигналом нейрона Y .

Таким образом, если обозначить функцию активации как F , выходной сигнал Y нейрона зависит от вектора входных значений $[X_1, X_2, \dots, X_n]$ и вектора весовых коэффициентов $[W_1, W_2, \dots, W_n]$ следующим образом:

$$Y = F(X_1 * W_1 + X_2 * W_2 + \dots + X_n * W_n)$$

На практике используются разные функции активации, позволяющие изменить поведение нейрона, например, линейная, пороговая, сигмоидная и др. Рассмотрим кратко для примера некоторые из этих функций. Более полное описание Вы найдете, например, в [12]

Линейная функция активации

График линейной функции активации показан на рис. 4-3. Как видите, этот график представляет собой прямую линию, угол наклона которой зависит от константы k , а величина сдвига по горизонтальной оси координат — от константы t .

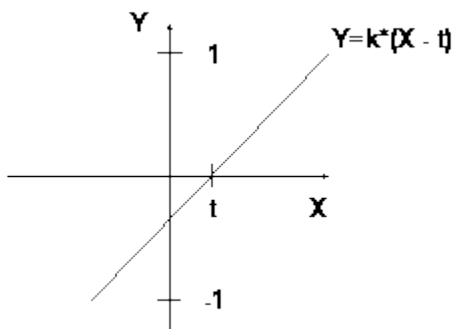


Рис. 4-3. Линейная функция активации

При использовании линейной функции активации значение на выходе нейрона равно взвешенной сумме, вычисляемой по следующей формуле:

$$Y = k * (X_1 * W_1 + X_2 * W_2 + \dots + X_n * W_n - t)$$

Константа t называется порогом нейрона [12]. Изменяя порог нейрона, можно сдвигать функцию активации по оси абсцисс.

Пороговая функция активации

Пороговая функция активации меняет свое значение скачкообразно, когда аргумент функции достигает значения порога нейрона t .

Используют два варианта пороговой функции — пороговую биполярную функцию активации и пороговую бинарную функцию активации.

На рис. 4-4 мы привели график пороговой биполярной функции, изменяющей свое значение от -1 до 1 .

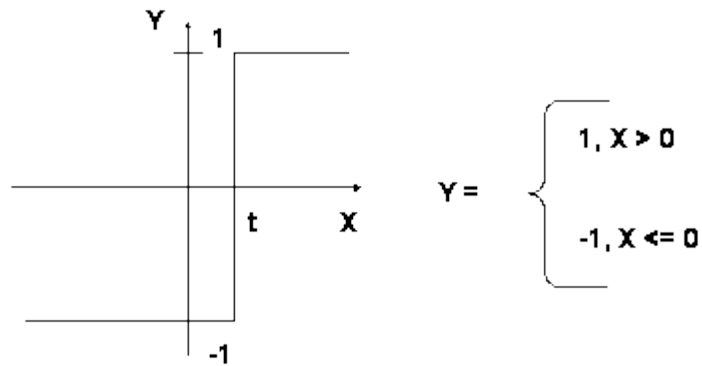


Рис. 4-4. Пороговая биполярная функция активации

Помимо биполярной пороговой функции активации применяется так называемая бинарная пороговая функция активации. Она может принимать значения 0 или 1, как это показано на рис. 4-5.

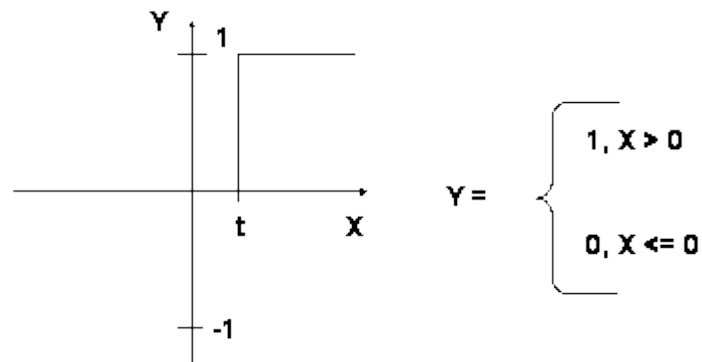


Рис. 4-5. Пороговая бинарная функция активации

Сигмоидная функция

Сигмоидная функция изменяет свое значение в диапазоне от 0 до 1, как это показано на рис. 4-6.

Особенностью сигмоидной функции является ее монотонность и дифференцируемость во всем диапазоне значений аргумента. Благодаря этому сигмоидная функция часто применяется при создании нейронных сетей.

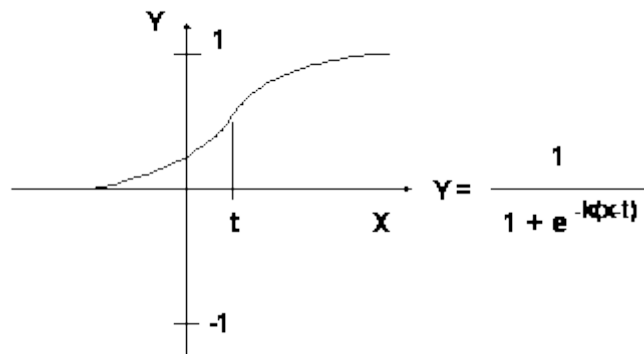


Рис. 4-6. Сигмоидная функция активации

Существует и так называемая биполярная сигмоидная функция активации, принимающая значения в диапазоне от -1 до 1 (рис. 4-7).

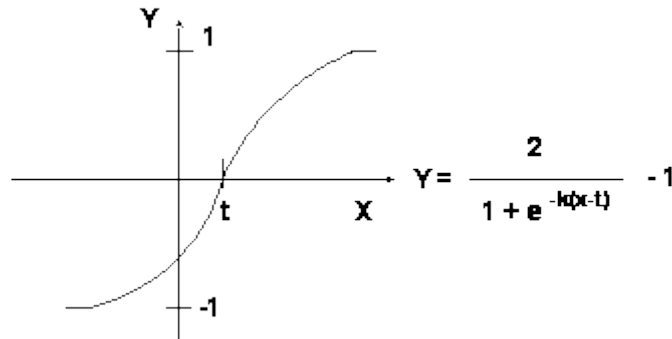


Рис. 4-7. Сигмоидная биполярная функция активации

Схематическое изображение нейрона

Как видите, модель нейрона достаточно проста. Существует немало реализаций этой модели, как аппаратных, так и программных. Аппаратные реализации модели нейрона хорошо описаны в [13]. Там же Вы найдете и ссылки на программные реализации искусственного нейрона.

Итак, рассмотренный выше искусственный нейрон можно представить себе как совокупность следующих компонентов:

- ☐ набор входов $[X_1, X_2, \dots, X_n]$, на которые попадают входные сигналы;
- ☐ вектор весовых коэффициентов $[W_1, W_2, \dots, W_n]$, определяющих реакцию нейрона на возбуждение от входных сигналов;
- ☐ сумматор, формирующий из всех входных сигналов взвешенную сумму с использованием коэффициентов $[W_1, W_2, \dots, W_n]$;
- ☐ функцию активации, формирующую из взвешенной суммы входных сигналов выходное значение.

На рис. 4-2 была показана подробная схема искусственного нейрона, включающая в себя все перечисленные выше компоненты. Однако на схемах нейронных сетей чаще используется упрощенная схема нейрона, которую мы привели на рис. 4-8.

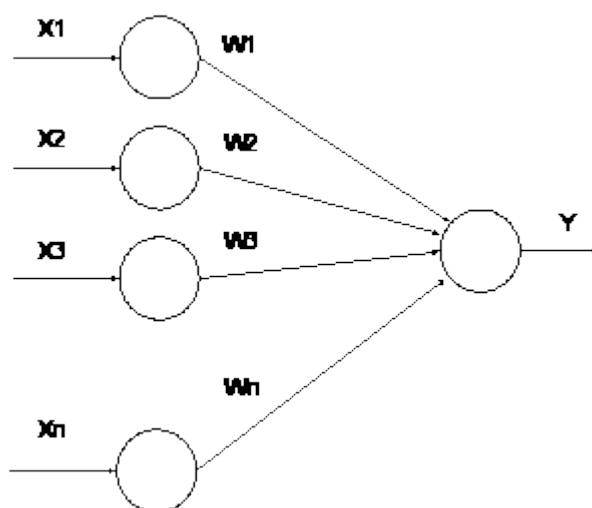


Рис. 4-8. Схематическое обозначение нейрона

Здесь входные значения $X_1 \dots X_n$ попадают на входные элементы (аналоги синапсов), обозначенные вертикальным рядом кружков. Влияние весовых коэффициентов отмечается на этой схеме при помощи набора значений $W_1 \dots W_n$.

Далее входные сигналы, умноженные на весовые коэффициенты, передаются в сумматор и функцию активации, для обозначения которых используется еще один кружок.

И, наконец, выход нейрона обозначен как Y .

Структуры нейронных сетей

Сам по себе искусственный нейрон представляет собой относительно простое устройство, выполняющее несложные функции. Из этих устройств можно формировать сети нейронов, объединяя отдельные нейроны связями.

Нейронным сетям по плечу решение достаточно сложных задач, таких, например, как распознавание образов и речи, кластеризация значений, предсказание значений сложных и неизвестных функций на основе анализа имеющихся значений этих функций и так далее.

Существует множество способов объединить отдельные нейроны в сеть. Чтобы успешно решить с помощью нейронной сети ту или иную задачу, необходимо правильно выбрать топологию сети, т.е. сколько в сети должно быть нейронов, и как они должны быть связаны между собой.

Кроме того, существует еще проблема обучения нейронных сетей, о которой мы расскажем позже в этой главе.

Однослойная нейронная сеть

Простейшая нейронная сеть состоит из одного слоя нейронов. Соответствующая топология нейронной сети показана на рис. 4-9.

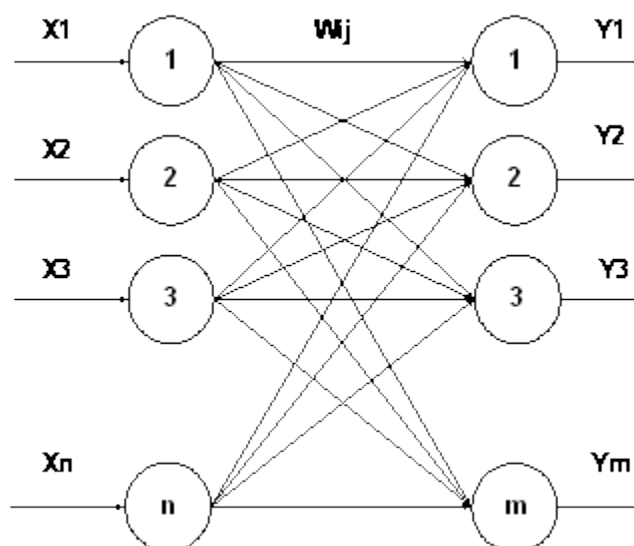


Рис. 4-9. Однослойная нейронная сеть

Слой нейронов — это такой набор нейронов, на которые в каждый момент времени параллельно поступает информация от других нейронных элементов сети [12]. Таким образом, элементы, обозначенные на рис. 4-9 цифрами от 1 до m представляют собой слой нейронов.

Первый слой нашей сети, на который подаются входные сигналы $x_1 \dots x_n$, играет роль распределительного слоя нейронной сети. Каждый элемент распределительного слоя нейронной сети имеет связи со всем нейронами следующего слоя, называемого обрабатывающим слоем.

Заметим, что хотя сеть фактически и состоит из двух слоев, одна называется однослойной. Первый распределительный слой не считается, так как он не занят обработкой информации, а играет вспомогательную роль.

Такую нейронную сеть можно представить себе в виде черного ящика, у которого имеется n входов и m выходов. Кроме этого, имеется набор (матрица) весовых коэффициентов, общее количество которых равно произведению n на m .

Примечательно то, что, изменяя матрицу весовых коэффициентов, можно полностью изменять поведение сети, т.е. ее реакцию на те или иные комбинации входных сигналов.

Забегая вперед, скажем, что в процессе обучения можно заставить сеть распознавать определенные комбинации сигналов и их значений. При этом один набор выходных значений может соответствовать одному набору входных значений, другой набор выходных значений — другому набору входных значений и так далее. Это свойство нейронной сети и позволяет применять ее для распознавания образов вообще, и для распознавания речи, в частности.

Персептрон Розенблатта

Другая структура нейронной сети, применяющаяся для распознавания образов, называется персептроном Розенблатта.

Такая сеть состоит из трех слоев, показанных на рис. 4-10.

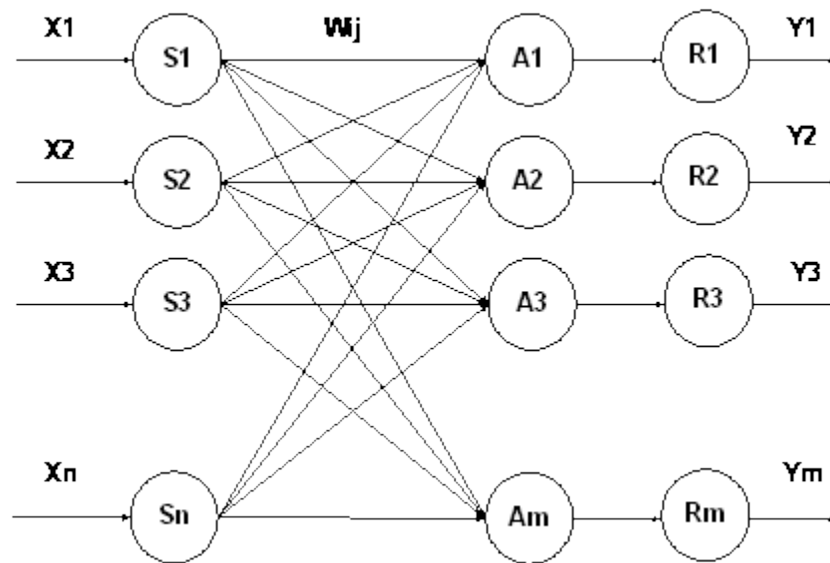


Рис. 4-10. Топология персептрона Розенблатта

Слой, обозначенный как $S_1 \dots S_n$, — это входной сенсорный слой. Его назначение состоит в том, чтобы воспринимать входные сигналы.

Слой $A_1 \dots A_m$ называется ассоциативным. Именно здесь происходит непосредственная обработка информации.

Что же касается слоя $R_1 \dots R_m$, то он называется эффекторным, и служит для передачи выходных воздействий. Особенностью этого слоя является использование в нейронах пороговой функции активации.

Многослойные нейронные сети

При исследовании мозга было обнаружено наличие в нем слоистых структур. С помощью искусственных нейронов можно моделировать такие структуры, создавая многослойные нейронные сети.

Структура такой многослойной сети показана на рис. 4-11.

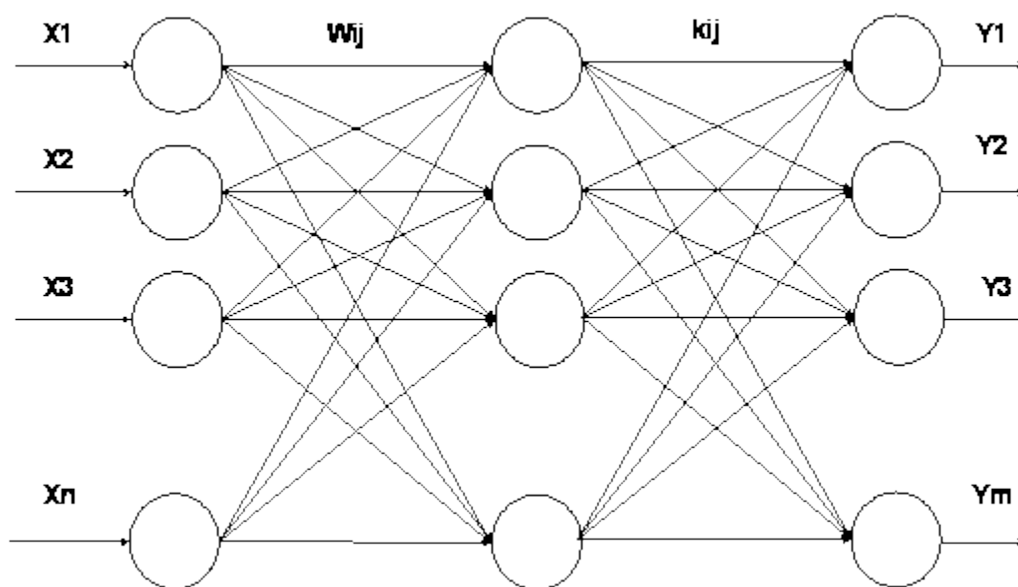


Рис. 4-11. Многослойная нейронная сеть

В каждом слое может находиться разное количество нейронов. При этом все входы нейронов следующего слоя соединены с выходами всех нейронов предыдущего слоя.

Все нейроны сети имеют свои наборы весовых коэффициентов. При моделировании весовые коэффициенты нейрона можно представить в виде массива. Аналогично, весовые коэффициенты всех нейронов одного слоя можно объединить в массив массивов.

Многослойные нейронные сети могут решать более сложные задачи, нежели однослойные. При этом, однако, промежуточные слои должны использовать нелинейные функции активации, например, сигмоидную функцию (рис. 4-6) или сигмоидную биполярную функцию (рис. 4-7).

Нейронные сети с обратными связями

Помимо рассмотренных выше «прямоточных» нейронных сетей, в которых сигналы распространяются в направлении от входов к выходам, существуют и более сложные структуры искусственных нейронов.

На рис. 4-12 мы показали структуру нейронной сети с обратными связями. Как видно на этом рисунке, сигналы с выходов сети попадают обратно на ее входы. При этом выходные сигналы смешиваются с входными сигналами.

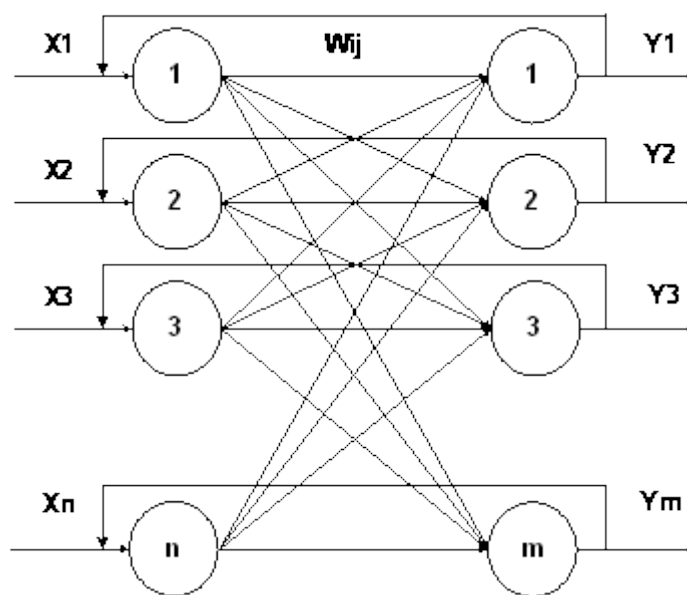


Рис. 4-12. Нейронная сеть с обратными связями

Если в нейронной сети имеются обратные связи, она может стать неустойчивой. Такая сеть может «зациклиться», проходя бесконечное количество раз через одни и те же состояния.

Нейронные сети могут быть усложнены за счет добавления в нее элементов, задерживающих распространение сигнала в прямом или обратном направлении. На рис. 4-13 показаны такие элементы задержки.

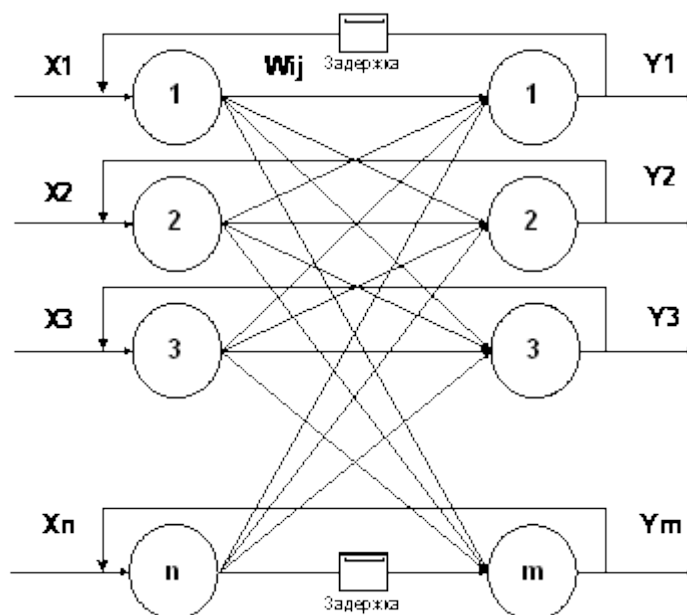


Рис. 4-13. Нейронная сеть с обратными связями и задержками

Используя элементы задержки при моделировании нейронной системы головного мозга, можно учитывать тот факт, что сигналы возбуждения передаются между нейронами не мгновенно.

Обучение нейронных сетей

Мы отмечали, что нейронную сеть можно представить себе в виде черного ящика, имеющего входы и выходы. Реакция нейронной сети на входные сигналы определяется ее структурой, а также весовыми коэффициентами, задающими «силу» связей между нейронами.

Сама по себе задача моделирования нейронных сетей при помощи компьютерных программ не представляет особой сложности. В Интернете можно найти немало программ моделирования нейронных сетей (попробуйте поискать по ключевым словам «Neural networks» и «нейронные сети»). Некоторые из них бесплатны, другие стоят довольно дорого, до тысяч долларов. Если Вы знакомы с программированием, то без особого труда сможете смоделировать такую сеть самостоятельно, например, на языках программирования C++, C# или других объектно-ориентированных языках.

Но вот выбор структуры нейронных сетей и выбор значений весовых коэффициентов для нейронов представляет собой две действительно сложные задачи.

Что касается структуры нейронной сети, то в зависимости от предназначения этой сети можно выбирать однослойную или многослойную сеть, а также вносить в нее дополнительные усложнения, вроде задержек. При этом современные исследователи имеют возможность опираться на предшествующий опыт работ в области нейронных сетей, так как большинство задач, вроде распознавания образов или прогнозирования, уже были детально изучены ранее. Например, для распознавания образов можно использовать нейронную сеть в виде персептрона Розенблатта (рис. 4-10). Другие задачи могут быть успешно решены при помощи нейронных сетей с другой структурой.

Выбрав структуру нейронной сети, нужно решить вторую проблему — выбрать значения весовых коэффициентов для всех нейронов этой сети. Значения этих коэффициентов определяются итерационно в процессе *обучения нейронной сети*.

Важно, что программист (или пользователь программы) никогда сам не задает весовые коэффициенты. Значения этих коэффициентов вычисляются в процессе обучения нейронной сети решению той или иной задачи. При этом одну и ту же нейронную сеть можно обучить решению разных задач, не внося никаких изменений в ее внутреннюю структуру.

Здесь может сразу возникнуть аналогия с обучением человека. Надо, однако, заметить, что сложность искусственных нейронных сетей не сопоставима со сложностью человеческого мозга. Кроме того, искусственные нейронные сети моделируют особенности живых нейронов лишь приблизительно. Поэтому при обсуждении способности искусственных нейронных сетей к обучению речь может идти только о выполнении простейших функций мозга. До реального

моделирования способностей человеческого мозга к обучению, к сожалению (или к счастью?), дело еще не дошло.

В литературе достаточно подробно описаны различные методики обучения нейронных сетей. Как правило, для того чтобы детально разобраться в этих методиках, необходимо вспомнить некоторые сведения из высшей математики, в частности, дифференцирование функций.

В нашей книге мы не будем утомлять читателя математическими уравнениями, а постараемся описать процесс обучения нейронных сетей без формул, на качественном уровне. Более подробную информацию и все необходимые математические выкладки Вы найдете в [11]-[15].

Задача распознавания объектов

Наилучшим образом искусственные нейронные сети подходят для решения задачи распознавания объектов, или, другими словами, *классификации объектов*.

Исходные данные для этой задачи — набор входных объектов, характеризующихся своими параметрами.

Возьмем такой практический пример. Пусть нам нужно создать систему, способную автоматически распознавать больных и здоровых людей по результатам измерения физиологических параметров (температура тела, нижнее и верхнее давление, частота сердцебиения), а также результатам анализа крови.

Простейший способ определить, здоров человек или болен, известен каждому. Берем градусник, измеряем температуру, и если она выше 37°C , считаем, что человек болен.

Разумеется, измеряя одну только температуру невозможно выявить все болезни. Например, если у человека слишком высокое или слишком низкое давление, температура тела может быть и нормальной. Тем не менее, такой человек болен.

Давайте попытаемся учесть давление. Будем считать здоровыми тех людей, у которых отношение верхнего давления к нижнему давлению находится вблизи значения $3/2$. Разумеется, это тоже очень сильное упрощение, однако оно дает нам возможность выявить больных людей с нормальной температурой тела.

На рис. 4-14 мы изобразили систему координат, где по горизонтали отсчитывается температура тела, а по вертикали — отношение верхнего давления к нижнему давлению.

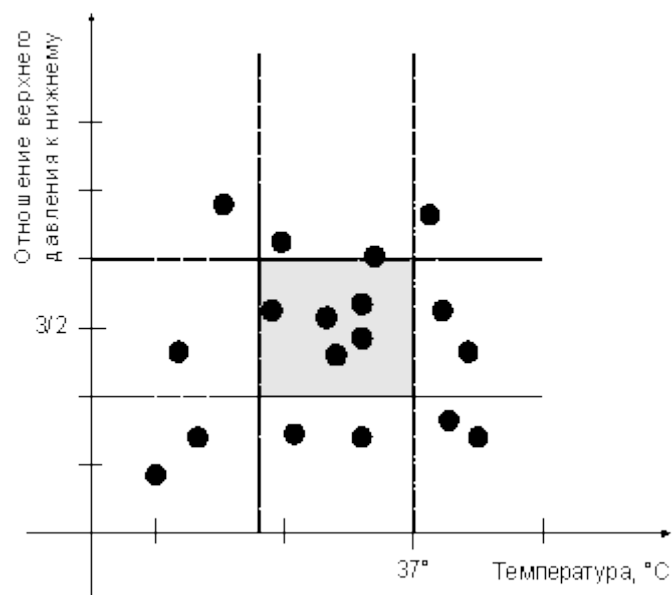


Рис. 4-14. Классификация по температуре и давлению

Точками здесь отмечаются пары значений температуры и отношения давлений для разных людей. Если точка попала в прямоугольную область, закрашенную серым цветом, то она соответствует здоровому человеку, если нет — больному.

Итак, мы решили простейшую задачу классификации объектов по двум параметрам, и представили на рис. 4-14 это решение в графическом виде.

Более точная классификация могла бы учитывать не отношение верхнего и нижнего давления, а абсолютные значения этих параметров. На рис. 4-15 мы привели классификацию по трем параметрам — верхнему давлению, нижнему давлению и температуре тела.

Теперь наш рисунок стал трехмерным. Если точка, соответствующая набору параметров человека, попала внутрь параллелепипеда, то этот человек здоров, если нет — болен.

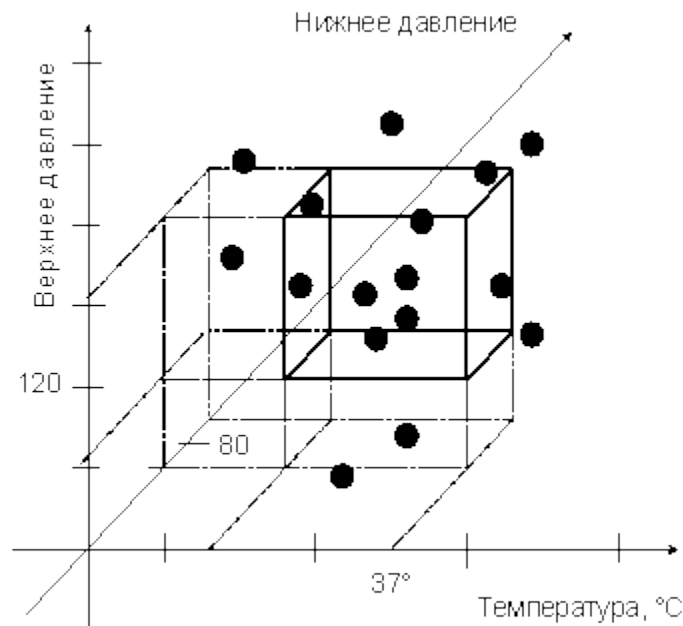


Рис. 4-15. Классификация по трем параметрам

Добавив в классификацию результаты анализов крови, мы превратим наше пространство параметров в многомерное пространство. Для разделения групп точек, расположенных в этом пространстве, нам потребуются многомерные гиперплоскости, которые невозможно изобразить на рисунке.

Искусственные нейронные сети в состоянии выполнить классификацию объектов по многим параметрам. Это можно сделать, например, с помощью персептрона Розенблатта.

Заметим, что персептрону все равно, какие объекты он распознает. Любые объекты представляются для него в виде точек, расположенных в некотором многомерном пространстве параметров. Персептрон группирует эти объекты в многомерные области, выполняя таким способом распознавание объектов.

Обучение персептрона

Хотя наша книга посвящена распознаванию и синтезу речи, а не распознаванию изображений, мы расскажем об обучении персептрона Розенблатта именно на примере распознавания изображений десятичных цифр. Как мы уже говорили, персептрон может распознавать любые объекты, представленные набором своих числовых параметров. Речевые сигналы также можно представить в виде таких параметров.

Итак, рассмотрим схему распознавания изображений десятичных цифр, которая есть во многих книгах, посвященных искусственным нейронным сетям (рис. 4-16).

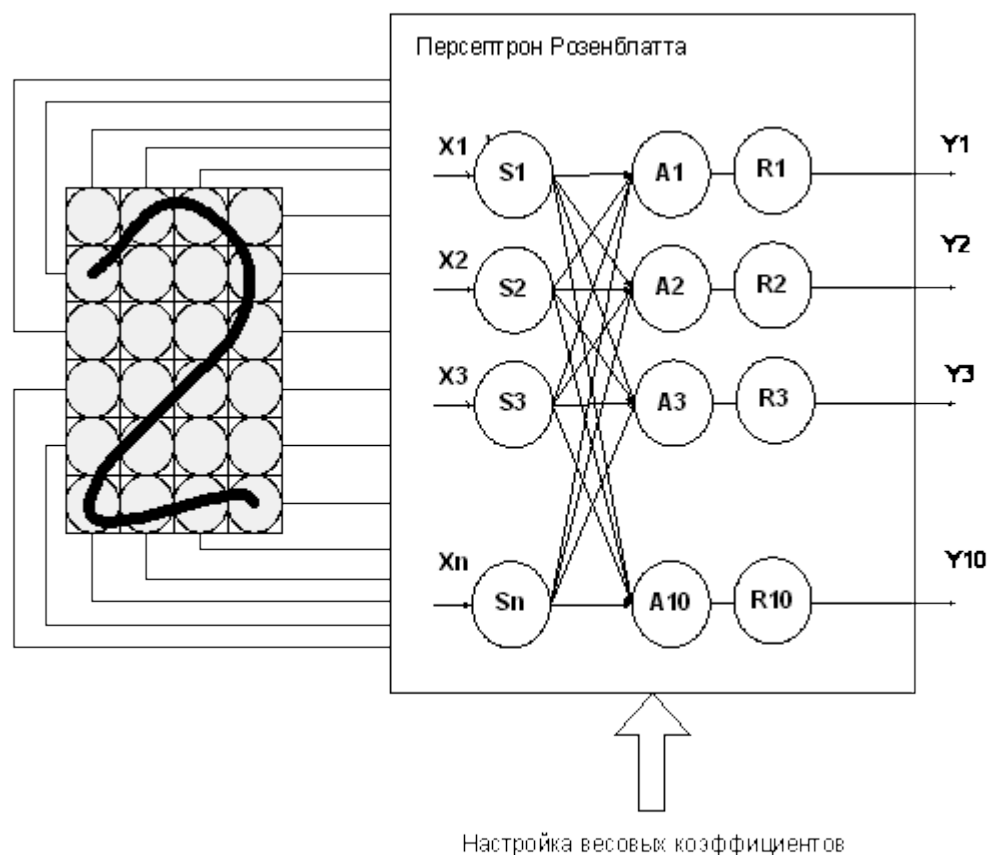


Рис. 4-16. Использование персептрона для распознавания цифр

В левой части рис. 4-16 располагается панель светочувствительных фотоэлементов. На эту панель можно класть прозрачные пластинки с изображением цифр от 0 до 9. Это изображение перекрывает часть поверхности некоторых фотоэлементов, из-за чего на них попадает меньше света.

Выход каждого фотоэлемента подключается к своему входу персептрона при помощи индивидуальной соединительной линии. Таким образом, количество n входных сенсорных элементов нашего персептрона равно количеству фотоэлементов, расположенных на панели.

Что же касается ассоциативного и эффекторного слоев, то в каждом из них имеется по 10 нейронов. В процессе обучения нейронной сети каждый из нейронов ассоциативного слоя настраивается таким образом, чтобы он реагировал на изображение «своей» цифры. Настройка происходит путем изменения весовых коэффициентов нейрона.

Как выполняется настройка?

Вначале на панель фотоэлементов кладут пластинку с изображением цифры 1. Затем программа настраивает весовые коэффициенты нейронной сети таким образом, чтобы сигнал появился только на одном выходе персептрона, например, на выходе Y_1 (и только на этом выходе). Далее на панель кладется пластинка с изображением цифры 2, а коэффициенты настраиваются, пока сигнал не появится только на выходе Y_2 . Затем эта операция повторяется для всех остальных цифр.

Рассмотрим процесс настройки весовых коэффициентов подробнее.

На первом этапе в качестве начальных значений весовых коэффициентов используются случайные числа. Далее персептрону предъявляется цифра 1, и проверяются значения сигналов на выходе эффекторного слоя $Y_1 \dots Y_{10}$.

Если значение сигнала на выходе Y_1 , который соответствует цифре 1, правильный, то весовые коэффициенты соответствующего нейрона ассоциативного слоя не меняются.

В том случае, если значение выходного сигнала неверно, то нужно изменить весовые коэффициенты нейрона, формирующего неправильный результат. При этом они немного уменьшаются или увеличиваются.

Далее снова проверяется значение сигнала на входе Y_1 , и, при необходимости, снова выполняется корректировка весовых коэффициентов. Эта процедура выполняется до тех пор, пока нейрон A_1 ассоциативного слоя не научится распознавать цифру 1. Затем персептрон обучается распознаванию других цифр.

Что же касается величины, на которую нужно изменять весовые коэффициенты в процессе обучения, то для его определения имеются различные готовые методы и формулы. Можно также выполнять прямой перебор всех значений, хотя это решение может отнять слишком много времени.

Способ определения изменений весовых коэффициентов в процессе обучения основан на оценке величины ошибки между реальным выходом нейронной сети, и тем выходом, который должен получиться. При этом вводится такое понятие, как *коэффициент обучения*. Этот коэффициент влияет на величину изменения весовых коэффициентов в процессе обучения, поэтому от него в широких пределах зависит скорость обучения нейронной сети.

Проблемы с обучением нейронных сетей

Как Вы уже, наверное, догадываетесь, проблема настройки весовых коэффициентов нейронной сети в процессе обучения не слишком проста. Лишь в конце 80-х годов был разработан способ обучения многослойных нейронных сетей. При этом было придумано правило, называемое *алгоритмом обратного распространения ошибок*.

Суть алгоритма обратного распространения ошибок состоит в выделении двух потоков нейронной сети. Прямой поток представляет собой поток сигналов, распространяющихся от входа сети к ее выходу. Обратный поток продвигается в направлении от выхода сети к ее входу и представляет собой поток ошибок. В соответствии с этими ошибками и происходит корректировка весовых коэффициентов слоев нейронной сети.

Не всякий процесс обучения нейронной сети будет успешным, или, как еще говорят, сойдется. Даже если обучение сети было успешным, не обязательно сеть будет выполнять свою работу наилучшим образом.

Почему такое происходит?

Во многих случаях задача обучения нейронной сети сводится к нахождению экстремума некоторой многомерной функции от нескольких параметров. При этом у одной и той же функции может быть несколько таких экстремумов.

Сказанное иллюстрирует рис. 4-17.

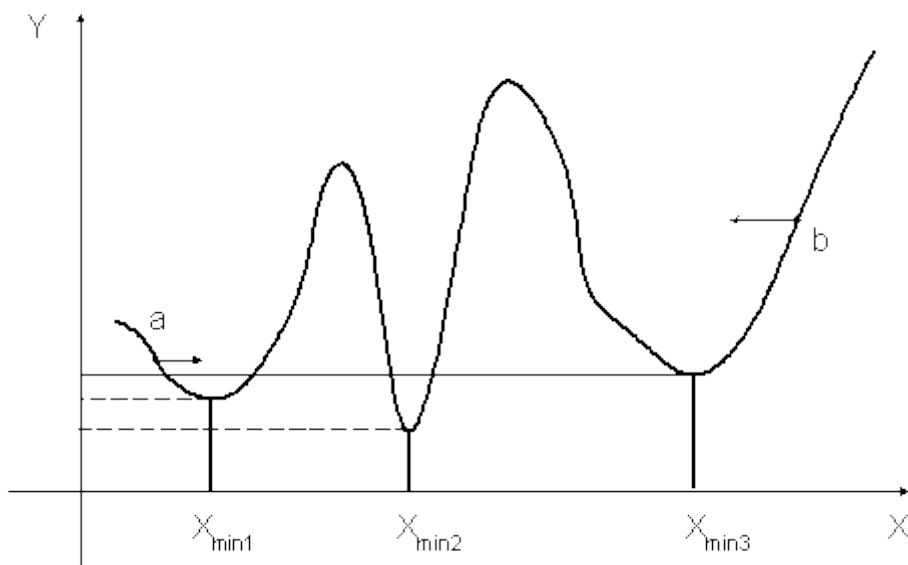


Рис. 4-17. Поиск экстремума функции

На этом рисунке изображена функция, имеющая три минимума X_{min1} , X_{min2} и X_{min3} . При этом только в точке X_{min2} достигается действительно наименьшее значение функции во всей области ее определения. Это глобальный минимум функции. Что же касается точек X_{min1} и X_{min3} , то это локальные минимумы функции.

Если поиск экстремума выполняется от точки, обозначенной на рис. 4-17 буквой a , то при движении по графику вправо будет обнаружен локальный минимум X_{min1} . Если же при поиске экстремума движение будет происходить из точки b влево, то будет обнаружен локальный минимум X_{min3} . В любом из этих случаев глобальный минимум функции, который находится в точке X_{min2} , то он так и не будет обнаружен. В [11] упомянут способ решения проблемы локальных экстремумов.

Существуют и другие проблемы, такие как слишком большие значения весовых коэффициентов, вызывающие «паралич» нейронной сети или чрезмерное уменьшение шага обучения, увеличивающее время обучения до чрезмерно большого значения. Кроме того, при некоторых условиях обучение нейронной сети может превратиться в неустойчивый процесс, когда, выучивая одно, сеть при этом забывает другое [11].

Таким образом, выбор метода обучения и коэффициента обучения, влияющего на скорость обучения, может оказаться критическим для успеха всего мероприятия. В любом случае применение на практике нейронных сетей нужно

рассматривать скорее как экспериментальные исследования, а не как решение простой и хорошо изученной инженерной задачи.

Обучение с учителем и без учителя

Существуют два метода обучения нейронных сетей — обучение с учителем и обучение без учителя.

Рассмотренная выше процедура обучения персептрона распознаванию цифр является процедурой с учителем. Учитель предъявляет сети по очереди все цифры от 0 до 9, обучая ее распознавать каждую цифру.

Но существует еще и другая методика обучения, без учителя. В этом случае в процессе обучения происходит самоорганизация сети. Сеть, обученная без учителя, сможет выполнять кластеризацию объектов, группируя их по признакам, определяющим свойства таких объектов.

Детерминистский и стохастический метод

Выше мы рассматривали процедуру обучения персептрона, представляющую собой последовательный пошаговый процесс. Такой метод обучения называется *детерминистским*.

В отличие от детерминистского метода, *стохастический* метод обучения предполагает псевдослучайное изменение весовых коэффициентов. При этом те изменения, которые приводят к уменьшению ошибки распознавания, сохраняются, а те изменения, которые ведут к увеличению ошибки, — отвергаются.

Стохастический метод обучения обладает некоторыми преимуществами. В частности, он позволяет решить проблему нахождения глобальных минимумов при наличии локальных минимумов. Стохастический метод соединяется с методом обратного распространения, что позволяет избавиться от некоторых недостатков метода обратного распространения (такими как проблемы с локальными минимумами и «паралич» сети).

Эксперименты с персептроном

Возможно, Вы уже устали от теории и хотели бы испытать нейронные сети в работе. В этом разделе мы расскажем Вам об очень интересной бесплатной реализации персептрона, созданной Константином Букреевым (konstantin@mail.primorye.ru).

Программа Константина Букреева называется GestureApp, и ее можно загрузить из каталогов программ Интернета по адресам www.codeproject.com или www.codeguru.com вместе с описанием и исходными текстами. Исходные тексты этой программы составлены на языке C++ для системы разработки приложений Microsoft Visual Studio 6.0.

Данная программа способна распознавать фигуры, показанные на рис. 4-18.



Рис. 4-18. Набор распознаваемых фигур

Нейронная сеть, использованная в программе GestureApp, содержит 32 нейрона и 29 аксонов (т.е. 29 эффекторов в терминах перцептрона Розенблатта). В нейронах ассоциативного слоя использована логарифмическая сигмоидальная функция активации. Для обучения сети применяется стандартный метод обратного распространения.

Конечно, программе GestureApp далеко до современных систем компьютерного зрения, способных различать лица в толпе. Но она поможет Вам почувствовать мощь технологии нейронных сетей. Что же касается программ распознавания речи, то о них мы начнем рассказывать в следующей главе.

6-7-ЛЕКЦИИ. МЕТОДЫ РАСПОЗНАВАНИЯ РЕЧИ. РАСПОЗНАВАНИЕ ПО ОБРАЗЦУ

План:

1. Методы распознавания речи.
2. Распознавание по образцу

Хотя термин «распознавания речи» встречается в литературе очень часто, на самом деле он имеет много различных значений. В этой главе мы попытаемся рассказать о различных трактовках этого термина, а также сделаем небольшой обзор технологий распознавания речи, сведения о которых доступны широкой публике.

Как правило, в существующих системах используются два принципиально разных подхода:

- ☐ распознавание голосовых меток;
- ☐ распознавание лексических элементов

Первый подход предполагает распознавание фрагментов речи по заранее записанному образцу. Этот подход широко используется в относительно простых системах, предназначенных для исполнения заранее записанных речевых команд.

Второй подход сложнее. При его реализации из потока речи выделяются отдельные лексические элементы — фонемы и аллофоны, которые затем объединяются в слоги и морфемы. Строго говоря, именно этот подход и используется в «настоящих» системах распознавания речи.

Все системы распознавания речи можно разделить на два класса:

- ☐ системы, зависящие от диктора;
- ☐ системы, не зависящие от диктора

К первому классу относятся системы, работа которых не зависит от диктора. Такие системы не требуют предварительного обучения и способны распознавать речь любого диктора.

Системы второго класса настраиваются на речь диктора в процессе обучения. Для работы с другим диктором такие системы требуют полной перенастройки.

Сразу отметим, что создание систем распознавания речи любого класса, пригодных для промышленного применения, представляет собой чрезвычайно сложную задачу. Как правило, разработчики таких систем обладают многолетним опытом в практическом применении речевых технологий.

Распознавание по образцу

Если у Вас есть современный мобильный телефон, то, скорее всего, одна из систем распознавания речи уже лежит в Вашем кармане. Такая система предназначена для ускоренного выбора абонентов из записной книжки мобильного телефона с помощью голоса.

Как это работает?

При добавлении нового контакта в записную книжку Вам предоставляется возможность ввести голосовую метку, идентифицирующую этот контакт, например, произнести в телефон имя или фамилию абонента. Возможно, придётся сделать это два или три раза.

Теперь, чтобы позвонить абоненту, достаточно нажать одну из кнопок, расположенных на корпусе мобильного телефона, и произнести голосовую метку. Номер абонента будет выбран из записной книжки, после чего мобильный телефон попытается установить с абонентом связь.

Помимо мобильного телефона, существуют и другие устройства с подобным голосовым управлением, например, компьютерные клавиатуры. Такие клавиатуры оборудуются встроенным микрофоном и позволяют назначить голосовые метки определенным клавишам, комбинациям клавиш или последовательностям клавиш. Разработчики клавиатур с голосовым интерфейсом утверждают, что с его помощью можно ускорить ввод информации,

однако качество распознавания голосовых меток зависит, например, от общего уровня шума в комнате.

Технология распознавания фрагментов по заранее записанным образцам применяется и во многих программах, позволяющих подключить голосовое управление к операционной системе Microsoft Windows и ее приложениям. При использовании этих программ Вы сможете запускать приложения, переключаться между ними, выбирать строки из меню и щелкать кнопки диалоговых окон, отдавая голосовые команды и не притрагиваясь руками к клавиатуре или мыши. Возможно, такие программы и не намного ускорят работу с приложениями для обычных людей, но они отчасти помогут инвалидам, неспособным использовать стандартные средства общения с компьютером.

Эта технология работает достаточно хорошо, если телефоном пользуется только один человек, а общее количество голосовых меток не превышает десятков-другой. Если Вы «обучите» свой телефон (или клавиатуру с голосовым интерфейсом) реагировать на Ваш голос, то только Вы и сможете пользоваться речевыми метками. Таким образом, эти системы относятся к классу систем, зависящих от диктора. Впрочем, этот недостаток есть и у многих более совершенных систем распознавания речи, основанных на выделении из речи лексических элементов.

8-9-ЛЕКЦИИ. ВЫДЕЛЕНИЕ ЛЕКСИЧЕСКИХ ЭЛЕМЕНТОВ РЕЧИ

План:

1. Выделение лексических элементов речи.

Далее в этой главе мы сосредоточим свое внимание на подходе к созданию систем распознавания речи, основанном на выделении из речи лексических элементов, таких как фонемы и аллофоны.

Как Вы можете убедиться, проводя осциллографические исследования (описанные нами в 3 главе), невозможно выделить из речи фонемы и аллофоны, анализируя только форму огибающей звукового сигнала. Как отмечено в [3], нельзя ограничиться составлением базы данных из записей звуковых сигналов всех фонем, аллофонов и других лексических элементов для последующего сравнения формы сигналов в процессе распознавания. Здесь нужны более сложные методы.

Предварительная обработка звуковых сигналов

Перед тем как предпринимать попытки распознавания речи, нужно выполнить предварительную обработку речевого сигнала. В ходе этой обработки следует удалить шумы и посторонние сигналы, частотный спектр которых

находится вне спектра человеческой речи. Такую обработку можно выполнить при помощи аналоговых или цифровых полосовых фильтров, рассмотренных нами во 2 главе этой книги.

Отфильтрованный звуковой сигнал нужно оцифровать, выполнив аналого-цифровое преобразование. Этот этап обработки мы тоже уже обсуждали ранее.

Всю предварительную обработку звукового сигнала можно сделать при помощи стандартного звукового адаптера, установленного в компьютере. Дополнительная цифровая обработка звукового сигнала (например, частотная фильтрация) может выполняться центральным процессором компьютера. Таким образом, при использовании современных персональных компьютеров системы распознавания речи не требуют для своей работы какого-либо специального аппаратного обеспечения.

Важным этапом предварительной обработки входного сигнала является нормализация уровня сигнала. Это позволяет уменьшить погрешности распознавания, связанные с тем, что диктор может произносить слова с разным уровнем громкости.

Заметим, однако, что если входной звуковой сигнал имеет слишком малый уровень громкости, то после нормализации может появиться шум. Поэтому для успешной работы системы распознавания речи необходимо отрегулировать оптимальным образом чувствительность микрофона. Чрезмерная чувствительность может привести к нелинейным искажениям сигнала и, как следствие, к увеличению погрешности распознавания речи.

Выделение информативных признаков речевого сигнала

Как мы уже говорили, информации об амплитуде и форме огибающей речевого сигнала не достаточно для выделения из речи лексических элементов. В зависимости от различных обстоятельств форма огибающей речевого сигнала может меняться в широких пределах, что затрудняет задачу распознавания.

Для решения задачи распознавания необходимо выделить первичные признаки речи, которые будут использованы на последующих этапах процесса распознавания. Первичные признаки выделяются посредством анализа спектральных и динамических характеристик речевого сигнала.

Спектральное представление речи

В [3] для выделения информативных признаков речевого сигнала используется спектральное представление речи. При этом на первом этапе осуществляется получение частотного спектра речевого сигнала с помощью набора программных полосовых фильтров (выполняя так называемое *дискретное преобразование Фурье*).

На втором этапе выполняются преобразования полученного спектра речевого сигнала:

- логарифмическое изменение масштаба в пространстве амплитуд и частот;
- сглаживание спектра с целью выделения его огибающей;
- кепстральный анализ (cepstral analysis), т.е. обратное преобразование Фурье от логарифма прямого преобразования [16].

Как отмечено в [3], перечисленные выше преобразования позволяют учитывать такие особенности речевого сигнала, как понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Учёт динамики речи

Помимо спектральных характеристик, необходимо учитывать и динамические особенности речи. Для этого используют *дельта-параметры*, представляющие собой производные по времени от основных параметров.

При этом мы можем отслеживать не только изменение параметров речи, но и скорость их изменения.

10-11-ЛЕКЦИИ. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ЗВУКОВЫХ СИГНАЛОВ

План:

1. Предварительная обработка звуковых сигналов.

Перед тем как предпринимать попытки распознавания речи, нужно выполнить предварительную обработку речевого сигнала. В ходе этой обработки следует удалить шумы и посторонние сигналы, частотный спектр которых находится вне спектра человеческой речи. Такую обработку можно выполнить при помощи аналоговых или цифровых полосовых фильтров, рассмотренных нами во 2 главе этой книги.

Отфильтрованный звуковой сигнал нужно оцифровать, выполнив аналого-цифровое преобразование. Этот этап обработки мы тоже уже обсуждали ранее.

Всю предварительную обработку звукового сигнала можно сделать при помощи стандартного звукового адаптера, установленного в компьютере. Дополнительная цифровая обработка звукового сигнала (например, частотная фильтрация) может выполняться центральным процессором компьютера. Таким образом, при использовании современных персональных компьютеров системы распознавания речи не требуют для своей работы какого-либо специального аппаратного обеспечения.

Важным этапом предварительной обработки входного сигнала является нормализация уровня сигнала. Это позволяет уменьшить погрешности

распознавания, связанные с тем, что диктор может произносить слова с разным уровнем громкости.

Заметим, однако, что если входной звуковой сигнал имеет слишком малый уровень громкости, то после нормализации может появиться шум. Поэтому для успешной работы системы распознавания речи необходимо отрегулировать оптимальным образом чувствительность микрофона. Чрезмерная чувствительность может привести к нелинейным искажениям сигнала и, как следствие, к увеличению погрешности распознавания речи.

Выделение информативных признаков речевого сигнала

Как мы уже говорили, информации об амплитуде и форме огибающей речевого сигнала не достаточно для выделения из речи лексических элементов. В зависимости от различных обстоятельств форма огибающей речевого сигнала может меняться в широких пределах, что затрудняет задачу распознавания.

Для решения задачи распознавания необходимо выделить первичные признаки речи, которые будут использованы на последующих этапах процесса распознавания. Первичные признаки выделяются посредством анализа спектральных и динамических характеристик речевого сигнала.

Спектральное представление речи

В [3] для выделения информативных признаков речевого сигнала используется спектральное представление речи. При этом на первом этапе осуществляется получение частотного спектра речевого сигнала с помощью набора программных полосовых фильтров (выполняя так называемое *дискретное преобразование Фурье*).

На втором этапе выполняются преобразования полученного спектра речевого сигнала:

- ☐ логарифмическое изменение масштаба в пространстве амплитуд и частот;
- ☐ сглаживание спектра с целью выделения его огибающей;
- ☐ кепстральный анализ (cepstral analysis), т.е. обратное преобразование Фурье от логарифма прямого преобразования [16].

Как отмечено в [3], перечисленные выше преобразования позволяют учитывать такие особенности речевого сигнала, как понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Учёт динамики речи

Помимо спектральных характеристик, необходимо учитывать и динамические особенности речи. Для этого используют *дельта-параметры*, представляющие собой производные по времени от основных параметров.

При этом мы можем отслеживать не только изменение параметров речи, но и скорость их изменения.

Выделение фонем и аллофонов

В первой главе нашей книги мы рассказывали о таких лексических элементах речи, как фонемы и аллофоны. Для их выделения в [3] применяются нейронные сети и метод формирования нейронных ансамблей.

При этом обучение выделению примитивов речи (фонем и аллофонов) может заключаться в формировании нейронных ансамблей, ядра которых соответствуют наиболее частой форме каждого примитива [17].

Формирование нейронных ансамблей представляет собой процесс обучения нейронной сети без учителя, при котором происходит статистическая обработка всех сигналов, поступающих на вход нейронной сети. При этом формируются ансамбли, соответствующие наиболее часто встречающимся сигналам. Запоминание редких сигналов происходит позже и требует подключения *механизма внимания* или иного контроля высшего уровня.

12-ЛЕКЦИЯ. ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ РЕЧЕВОГО СИГНАЛА

План:

1. Выделени информативных признаков речевого сигнала.

Как мы уже говорили, информации об амплитуде и форме огибающей речевого сигнала не достаточно для выделения из речи лексических элементов. В зависимости от различных обстоятельств форма огибающей речевого сигнала может меняться в широких пределах, что затрудняет задачу распознавания.

Для решения задачи распознавания необходимо выделить первичные признаки речи, которые будут использованы на последующих этапах процесса распознавания. Первичные признаки выделяются посредством анализа спектральных и динамических характеристик речевого сигнала.

Спектральное представление речи

В [3] для выделения информативных признаков речевого сигнала используется спектральное представление речи. При этом на первом этапе осуществляется получение частотного спектра речевого сигнала с помощью набора программных полосовых фильтров (выполняя так называемое *дискретное преобразование Фурье*).

На втором этапе выполняются преобразования полученного спектра речевого сигнала:

- логарифмическое изменение масштаба в пространстве амплитуд и частот;
- сглаживание спектра с целью выделения его огибающей;
- кепстральный анализ (cepstral analysis), т.е. обратное преобразование Фурье от логарифма прямого преобразования [16].

Как отмечено в [3], перечисленные выше преобразования позволяют учитывать такие особенности речевого сигнала, как понижение информативности высокочастотных участков спектра, логарифмическую чувствительность человеческого уха, и т.д.

Учёт динамики речи

Помимо спектральных характеристик, необходимо учитывать и динамические особенности речи. Для этого используют *дельта-параметры*, представляющие собой производные по времени от основных параметров.

При этом мы можем отслеживать не только изменение параметров речи, но и скорость их изменения.

Выделение фонем и аллофонов

В первой главе нашей книги мы рассказывали о таких лексических элементах речи, как фонемы и аллофоны. Для их выделения в [3] применяются нейронные сети и метод формирования нейронных ансамблей.

При этом обучение выделению примитивов речи (фонем и аллофонов) может заключаться в формировании нейронных ансамблей, ядра которых соответствуют наиболее частой форме каждого примитива [17].

Формирование нейронных ансамблей представляет собой процесс обучения нейронной сети без учителя, при котором происходит статистическая обработка всех сигналов, поступающих на вход нейронной сети. При этом формируются ансамбли, соответствующие наиболее часто встречающимся сигналам. Запоминание редких сигналов происходит позже и требует подключения *механизма внимания* или иного контроля высшего уровня.

13-ЛЕКЦИЯ. ВЫДЕЛЕНИЕ ФОНЕМ И АЛЛОФОНОВ

План:

1. Выделение фонем и аллофонов.

В первой главе нашей книги мы рассказывали о таких лексических элементах речи, как фонемы и аллофоны. Для их выделения в [3] применяются нейронные сети и метод формирования нейронных ансамблей.

При этом обучение выделению примитивов речи (фонем и аллофонов) может заключаться в формировании нейронных ансамблей, ядра которых соответствуют наиболее частой форме каждого примитива [17].

Формирование нейронных ансамблей представляет собой процесс обучения нейронной сети без учителя, при котором происходит статистическая обработка всех сигналов, поступающих на вход нейронной сети. При этом формируются ансамбли, соответствующие наиболее часто встречающимся сигналам. Запоминание редких сигналов происходит позже и требует подключения *механизма внимания* или иного контроля высшего уровня.

Уровни распознавания слитной речи

Распознавание слитной речи представляет собой многоуровневый процесс. После предварительной обработки речевого сигнала и выделения из него информативных признаков выполняется выделение лексических элементов речи. Это первый уровень распознавания.

На втором уровне выделяются слоги и морфемы, на третьем — слова, предложения и сообщения (рис. 5-1).

Как отмечается в [3], на каждом уровне сигнал кодируется представителями предыдущих уровней. То есть слоги и морфемы состояются из фонем и аллофонов, слова — из слогов и морфем, предложения и сообщения — из слов.

При переходе с уровня на уровень помимо представителей сигналов передаются и некоторые дополнительные признаки, временные зависимости и отношения между сигналами. Собирая сигналы с предыдущих уровней, высшие уровни располагают большим объемом информации (или её другим представлением), и могут осуществлять управление процессами на низших уровнях, например, с привлечением *механизма внимания*.

Механизм внимания используется при обучении нейронной сети. В случае использования такого механизма при появлении образца, неизвестного нейронной сети, скорость обучения многократно возрастает. При этом редко встречающийся образец запоминается в нейронной сети.

Применение нейронных сетей для распознавания речи

Рассказывая в 4 главе нашей книги о нейронных сетях, мы упоминали такое свойство этих сетей, как способность к обучению и классификации объектов по их числовым параметрам.

При обучении сети с учителем можно научить сеть распознавать объекты, принадлежащие заранее определенному набору классов. Если же сеть обучается без учителя, то она может группировать объекты по классам в соответствии с их цифровыми параметрами.

Таким образом, на базе нейронных сетей можно создавать обучаемые и самообучающиеся системы. В [3] формулируются следующие требования к

самообучающимся системам. Ниже мы приводим эти требования, взятые из [3] практически без изменений.

- Разработка системы заключается только в построении архитектуры системы

В процессе создания системы разработчик создает только функциональную часть, но не наполняет (или наполняет в минимальных объемах) систему информацией. Основную часть информации система получает в процессе обучения.

- Возможность контроля своих действий с последующей коррекцией

Этот принцип говорит о необходимости обратной связи *Действие-Результат-Коррекция* в системе. Такие цепочки очень широко распространены в сложных биологических организмах и используются на всех уровнях — от контроля мышечных сокращений на самом низком уровне до управления сложными механизмами поведения.

- Возможность накопления знаний об объектах рабочей области

Знание об объекте — это способность манипулировать его образом в памяти.

Количество знаний об объекте определяется не только набором его свойств, но ещё и информацией о его взаимодействии с другими объектами, поведении при различных воздействиях, нахождении в разных состояниях, и т.д., т.е. его поведении во внешнем окружении.

Например, знание о геометрическом объекте предполагает возможность предсказать вид его перспективной проекции при любом повороте и освещении. Это свойство наделяет систему возможностью абстрагирования от реальных объектов, т.е. возможностью анализировать объект при его отсутствии, открывая тем самым новые возможности в обучении.

- Автономность системы

При интеграции комплекса действий, которые система способна совершать, с комплексом датчиков, позволяющих контролировать свои действия и внешнюю среду, наделенная вышеприведенными свойствами система будет способна взаимодействовать с внешним миром на довольно сложном уровне.

При этом она будет адекватно реагировать на изменение внешнего окружения (естественно, если это будет заложено в систему на этапе обучения). Способность корректировать свое поведение в зависимости от внешних условий позволит частично или полностью устранить необходимость контроля извне, т.е. система станет *автономной*.

Возможность создания на базе искусственных нейронных сетей самообучающихся систем является важной предпосылкой для их применения в системах распознавания (и синтеза) речи.

14-ЛЕКЦИЯ. УРОВНИ РАСПОЗНАВАНИЯ СЛИТНОЙ РЕЧИ

План:

1. Уровни распознавания слитной речи.

Распознавание слитной речи представляет собой многоуровневый процесс. После предварительной обработки речевого сигнала и выделения из него информативных признаков выполняется выделение лексических элементов речи. Это первый уровень распознавания.

На втором уровне выделяются слоги и морфемы, на третьем — слова, предложения и сообщения (рис. 5-1).

Как отмечается в [3], на каждом уровне сигнал кодируется представителями предыдущих уровней. То есть слоги и морфемы состояются из фонем и аллофонов, слова — из слогов и морфем, предложения и сообщения — из слов.

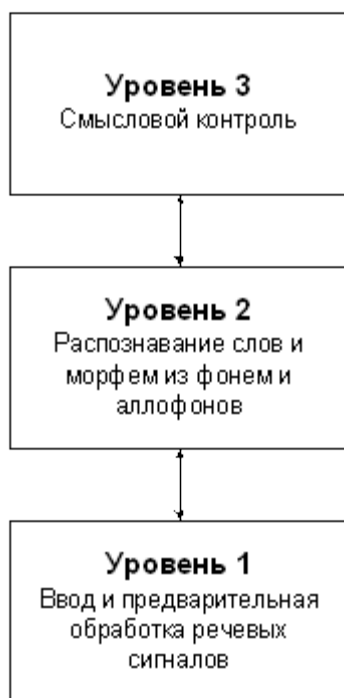


Рис. 5-1. Три уровня распознавания слитной речи

При переходе с уровня на уровень помимо представителей сигналов передаются и некоторые дополнительные признаки, временные зависимости и отношения между сигналами. Собирая сигналы с предыдущих уровней, высшие уровни располагают большим объемом информации (или её другим представлением), и могут осуществлять управление процессами на низших уровнях, например, с привлечением *механизма внимания*.

Механизм внимания используется при обучении нейронной сети. В случае использования такого механизма при появлении образца, неизвестного

нейронной сети, скорость обучения многократно возрастает. При этом редко встречающийся образец запоминается в нейронной сети.

15-ЛЕКЦИЯ. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ РЕЧИ

План:

1. Применение нейронных сетей для распознавания речи.

Рассказывая в 4 главе нашей книги о нейронных сетях, мы упоминали такое свойство этих сетей, как способность к обучению и классификации объектов по их числовым параметрам.

При обучении сети с учителем можно научить сеть распознавать объекты, принадлежащие заранее определенному набору классов. Если же сеть обучается без учителя, то она может группировать объекты по классам в соответствии с их цифровыми параметрами.

Таким образом, на базе нейронных сетей можно создавать обучаемые и самообучающиеся системы. В [3] формулируются следующие требования к самообучающимся системам. Ниже мы приводим эти требования, взятые из [3] практически без изменений.

- ☐ Разработка системы заключается только в построении архитектуры системы

В процессе создания системы разработчик создает только функциональную часть, но не наполняет (или наполняет в минимальных объемах) систему информацией. Основную часть информации система получает в процессе обучения.

- ☐ Возможность контроля своих действий с последующей коррекцией

Этот принцип говорит о необходимости обратной связи *Действие-Результат-Коррекция* в системе. Такие цепочки очень широко распространены в сложных биологических организмах и используются на всех уровнях — от контроля мышечных сокращений на самом низком уровне до управления сложными механизмами поведения.

- ☐ Возможность накопления знаний об объектах рабочей области

Знание об объекте — это способность манипулировать его образом в памяти.

Количество знаний об объекте определяется не только набором его свойств, но ещё и информацией о его взаимодействии с другими объектами, поведении при различных воздействиях, нахождении в разных состояниях, и т.д., т.е. его поведении во внешнем окружении.

Например, знание о геометрическом объекте предполагает возможность предсказать вид его перспективной проекции при любом повороте и освещении. Это свойство наделяет систему возможностью абстрагирования от реальных

объектов, т.е. возможностью анализировать объект при его отсутствии, открывая тем самым новые возможности в обучении.

□ Автономность системы

При интеграции комплекса действий, которые система способна совершать, с комплексом датчиков, позволяющих контролировать свои действия и внешнюю среду, наделенная вышеприведенными свойствами система будет способна взаимодействовать с внешним миром на довольно сложном уровне.

При этом она будет адекватно реагировать на изменение внешнего окружения (естественно, если это будет заложено в систему на этапе обучения). Способность корректировать свое поведение в зависимости от внешних условий позволит частично или полностью устранить необходимость контроля извне, т.е. система станет *автономной*.

Возможность создания на базе искусственных нейронных сетей самообучающихся систем является важной предпосылкой для их применения в системах распознавания (и синтеза) речи.

Представление речи в виде набора числовых параметров

После выделения информативных признаков речевого сигнала можно представить эти признаки в виде некоторого набора числовых параметров (т.е. в виде вектора в некотором числовом пространстве). Далее задача распознавания примитивов речи (фонем и аллофонов) сводится к их классификации при помощи обучаемой нейронной сети.

Нейронные сети можно использовать и более высоких уровнях распознавания слитной речи для выделения слогов, морфем и слов.

Нейронные ансамбли

В [3] отмечается, что в качестве модели нейронной сети, пригодной для распознавания речи и обучаемой без учителя можно выбрать самоорганизующуюся карту признаков Кохонена. В ней для множества входных сигналов формируется нейронные ансамбли, представляющие эти сигналы. Этот алгоритм обладает способностью к статистическому усреднению, что позволяет решить проблему изменчивости речи.

По сравнению с классическим программированием, когда алгоритм решения той или иной задачи задан жестко, нейронные сети позволяют динамически изменять алгоритм простым изменением архитектуры сети.

Генетические алгоритмы

Возможность изменения алгоритма работы нейронной сети простым изменением ее архитектуры позволяют решать задачи совершенно новым способом, с помощью так называемых *генетических алгоритмов*.

При использовании генетических алгоритмов создаются правила отбора, позволяющие определить, лучше или хуже справляется новая нейронная сеть с

решением задачи. Кроме того, определяются правила модификации нейронной сети.

Изменяя достаточно долго архитектуру нейронной сети и отбирая те архитектуры, которые позволяют решить задачу наилучшим образом, рано или поздно можно получить верное решение задачи.

Генетические алгоритмы обязаны своим появлением эволюционной теории (отсюда и характерные термины: популяция, гены, родители-потомки, скрещивание, мутация). Таким образом, существует возможность создания таких нейронных сетей, которые ранее не изучались исследователями (или не поддаются аналитическому изучению), но, тем не менее, успешно решают задачу [3].

16-ЛЕКЦИЯ. ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ ПРИ РАСПОЗНАВАНИИ РЕЧИ

План:

1. Применение Вейвлет-преобразований при распознавании речи.

В только что описанной системе SAS для выделения из речи синтаксических элементов применялось быстрое преобразование Фурье.

Однако, как отмечают исследователи, анализ Фурье обладает целым рядом недостатков, в результате которых происходит потеря информации о временных характеристиках обрабатываемых сигналов. Этот анализ подразумевает использование искусственных приемов, с помощью которых осуществляется частотно-временная локализация, например, окон данных (на рис. 5-3 это окно обозначено как **Окно ввода**).

В современных технологиях обработки и распознавания сигналов применяются так называемые вейвлет-преобразования и вейвлет-анализ.

Термин *вейвлет* (wavelets) можно перевести как «маленькая волна». Вейвлеты представляют собой новый инструмент решения различных задач прикладной математики. Вейвлет-анализ, детальное знакомство с которым требует определенных познаний в математике, лишен недостатков анализа Фурье. Он позволяет достичь неплохих результатов при использовании в системах распознавания речи.

В чем отличие анализа Фурье от вейвлет-анализа?

Фурье-анализ предполагает разложение исходной периодической функции в ряд, в результате чего исходная функция может быть представлена в виде суперпозиции синусоидальных волн различной частоты. Такая суперпозиция и есть спектр сигнала, о котором мы говорили в 3 главе нашей книги.

Что же касается вейвлет-анализа, то здесь входной сигнал раскладывается в базис функций, характеризующих как частоту, так и время. Поэтому с помощью вейвлетов можно анализировать свойства сигнала одновременно и в физическом

пространстве (время, координата), и в частотном пространстве. Чтобы подчеркнуть такое обстоятельство, в зарубежной литературе Фурье-анализ называют single spectrum, а спектры, полученные на основе вейвлет-преобразований — itime-scale spectrum, или wavelet spectrum [19].

Функции-базисы для вейвлетных преобразований конструируются на основе производных функций Гаусса. Подробнее об этом Вы сможете прочитать в [19].

На рис. 5-6, взятом из [19], показаны наиболее часто используемые вейвлеты.

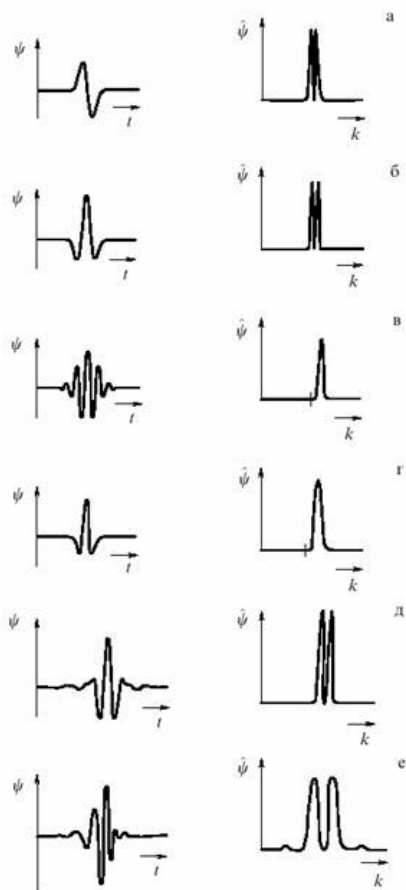


Рис. 5-6. Часто используемые вейвлеты

Эти функции имеют свои названия. Мы привели их в табл. 5-1.

Таблица 5-1. Часто используемые вейвлеты

| Обозначение на рис. 5-6 | Название |
|-------------------------|---|
| а | WAVE-вейвлет |
| б | МНАТ-вейвлет. Получил свое название от «мексиканская шляпа, сомbrero» (Mexican Hat) |
| в | Morlet |

| | |
|---|------------|
| г | Paul |
| д | LMB |
| е | Daubeshies |

При использовании вейвлет-преобразований для распознавания речи разработчик должен выбрать нужную функцию. От правильного выбора зависит успешность распознавания.

17-ЛЕКЦИЯ. СИСТЕМЫ ГОЛОСОВОГО УПРАВЛЕНИЯ

План:

1. Системы голосового управления.

Все системы голосового управления компьютером и другим оборудованием можно условно разделить на два класса.

К первому классу мы будем относить устройства, назначением которых является облегчение работы с компьютером. Такие устройства используются в качестве дополнительного канала управления наряду с основными и традиционными каналами — клавиатура и мышь, а также экран компьютера (который служит каналом обратной связи).

В устройствах второго класса речевой интерфейс является основным и единственным. Эти устройства доступны тем, кто не может работать с клавиатурой и экраном компьютера либо из-за проблем со здоровьем, либо из-за чрезмерной занятости какой-либо работой (например, управлением автомобилем или самолетом).

В этом разделе мы рассмотрим несколько систем голосового управления, созданных в компании «Центр речевых технологий», которую мы упоминали в 3 главе нашей книги. Это программы VoiceNavigator и Truffaldino, библиотека распознавания голосовых команд VoiceKeyKit и система DiVo управления мобильным телефоном в автомобиле без использования рук (hands-free car kit).

Все они, кроме последней, относятся к первому классу, т.к. для работы с ними необходимы традиционные средства ввода и вывода информации. Систему DiVo управления мобильным телефоном в автомобиле можно отнести ко второму классу — для работы с этой системой не нужен ни компьютер, ни экран, ни клавиатура.

Программа VoiceNavigator

Программа VoiceNavigator (рис. 6-1) является типичным представителем программ голосового управления компьютером. Она позволяет пользователю запускать приложения голосом, не дотрагиваясь до клавиатуры, и выполнять произвольно заданные команды.

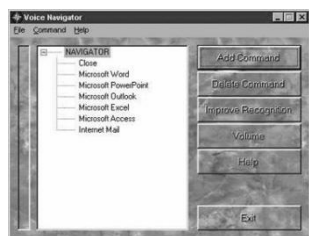


Рис. 6-1. Программа VoiceNavigator

Перед использованием программы VoiceNavigator ее необходимо обучить, произнеся в микрофон слова команд. Так как программа VoiceNavigator распознает команды по образцам, то команды можно произносить на любом языке и любым голосом.

Чтобы программа начала распознавать голосовые команды, ее необходимо «разбудить», произнеся ключевое слово. После этого программа будет реагировать только на Ваши команды, игнорируя остальные звуки.

В программе есть функция голосового ответа-подтверждения команд. Эта функция позволяет убедиться, что Ваша команда распознана системой и готова для выполнения.

Программа VoiceNavigator нетребовательна к ресурсам компьютера. Вы можете использовать ее в компьютере, оборудованном процессором с тактовой частотой 200 МГц или выше, причем для ввода звуковых команд подойдет любой звуковой адаптер, например, Creative Sound Blaster.

Программа Truffaldino

Программа Truffaldino (рис. 6-2) предназначена для управления домашним или офисным оборудованием, таким как телевизор, домофон, видео-магнитофон, аудио-система, кондиционер, микроволновая печь, система освещения, телефон и т.п.

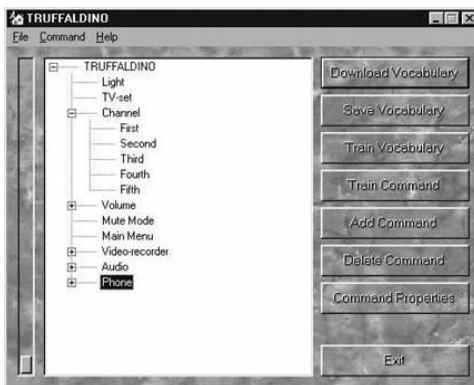


Рис. 6-2. Программа Truffaldino

Эта программа пригодится людям с ограниченными физическими возможностями, не способными управляться с домашней техникой обычным способом.

После предварительного обучения программа Truffaldino способна распознавать несколько десятков команд, произнесенных на любом языке и любым голосом. Последнее возможно благодаря тому, что команды распознаются по голосовым шаблонам.

Особенностью программы является хорошая работа в условиях шумов. Это обстоятельство имеет большое значение, так как дома или в офисе акустическая обстановка может быть далеко не идеальной.

Помимо ключевой фразы, предназначенной для «пробуждения» программы Truffaldino, имеется функция ограничения доступа. С помощью этой функции Вы сможете заблокировать запуск программы посторонними людьми.

В программе Truffaldino есть функция голосового ответа-подтверждения команд, аналогичная такой же функции программы VoiceNavigator. Эта функция позволяет убедиться, что Ваша команда распознана системой и готова для выполнения.

Системные требования программы Truffaldino аналогичны требованиям программы VoiceNavigator. Однако для управления оборудованием необходимо применение дополнительного устройства, такого как Infra-Red X-tender.

Библиотека распознавания голосовых команд VoiceCom

Библиотека распознавания голосовых команд VoiceCom составляет ядро описанных выше программ VoiceNavigator и Truffaldino. С ее помощью разработчики могут добавить голосовое управление в создаваемые ими приложения.

Разработка собственных средств голосового управления требует специальных знаний. Она может отнять у разработчиков приложений немало времени и сил. Однако воспользовавшись готовой библиотекой VoiceCom, разработчики могут легко добавить в приложения следующие функциональные возможности:

- ☐ управление оборудованием с помощью голоса;
- ☐ выполнение речевых запросов к базам данных через микрофон или даже по телефону;
- ☐ поиск по ключевым словам в звуковых WAV-файлах

Следует отметить, что библиотека VoiceCom позволяет встраивать голосовые функции не только в обычные программы для персональных компьютеров, но и в автономные устройства, оборудованные цифровыми сигнальными процессорами DSP.

Алгоритмы, реализованные в библиотеке распознавания голосовых команд VoiceCom, обладают высоким быстродействием, нетребовательны к объему оперативной памяти и способны адаптироваться к шумам.

Библиотека VoiceCom обеспечивает распознавание команд, произнесенных любым голосом и на любом языке. При этом имеется возможность структурирования для практически неограниченного словаря.

При этом алгоритмы позволяют распознавать 100-200 команд с предварительным обучением для каждого диктора, и 30-50 команд для любого диктора (в режиме, не зависящем от диктора). Если команды произносятся по телефону, то алгоритмы библиотеки VoiceCom позволяют распознать 10-20 слов, произнесенных любым диктором.

Ну и, конечно, в библиотеке реализована возможность активации распознавания команд по ключевому слову, исключающая неожиданные реакции системы на посторонние звуки.

Комплекс DiVo

Известно, что пользование мобильным телефоном во время вождения автомобиля может привести к возникновению аварийной ситуации на дороге.

В большинстве стран Европы, включая Россию, и в ряде штатов США, в целях обеспечения безопасности движения принят закон, требующий от водителей автомобилей использования специального оборудования, позволяющего разговаривать по мобильному телефону без использования рук (в режиме hands-free).

Надо заметить, что компании, выпускающие мобильные телефоны, комплектуют их таким оборудованием. Однако комплекс DiVo (рис. 6-3), разработанный в компании «Центр речевых технологий», значительно превосходит существующие системы по качеству распознавания речи в условиях интенсивных шумов и помех.

Этот комплекс позволяет полностью управлять мобильным телефоном стандарта GSM в салоне автомобиля посредством голосовых команд. Он активизируется при помощи ключевой команды «пробуждения» и позволяет пользоваться мобильным телефоном, не отвлекаясь от управления автомобилем.

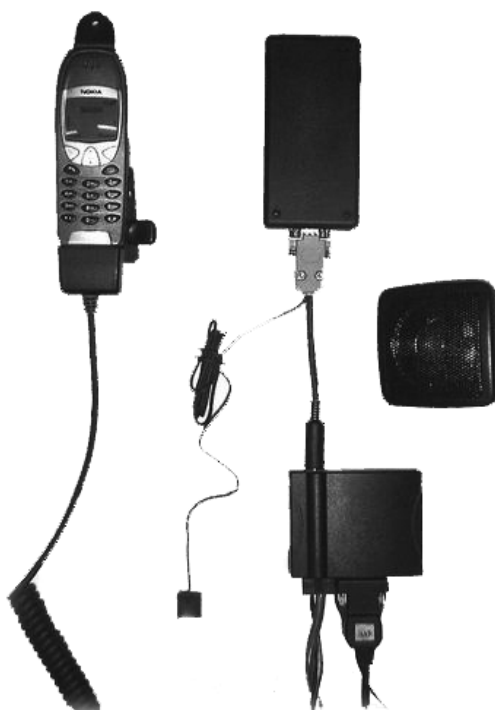


Рис. 6-3. Комплекс DiVo

Что входит в состав комплекса DiVo?

Комплект поставки DiVo показан на рис. 6-3. В него входит следующее оборудование:

- ☐ компактный центральный блок, выполняющий функции управления, обработки и распознавания речи;
- ☐ высококачественный направленный микрофон;
- ☐ держатель для мобильного телефона с кабелем питания и разъемом для прикуривателя (комплекс может питаться как от прикуривателя, так и непосредственно от электросети автомобиля);
- ☐ кнопка управления системой;
- ☐ компактный громкоговоритель;
- ☐ кабель отключения звука автомобильной аудио-системы;
- ☐ антенный кабель для подключения внешней антенны

Существующая версия комплекса DiVo может работать совместно с мобильными телефонами Ericsson, Nokia и Siemens.

Водитель может выдавать комплексу DiVo до 20 голосовых команд, таких как набор телефонного номера, повтор последнего набравшегося номера, выбор номера телефона из голосовой телефонной книги, емкость которой составляет 40-50 записей.

Для наполнения телефонной книги комплекса DiVo можно использовать содержимое телефонной книги, хранящейся в памяти мобильного телефона.

Так как в составе комплекса DiVo не предусмотрен монитор (им просто было бы невозможно пользоваться во время вождения автомобиля), все операции выполняются с помощью удобного и разветвленного голосового меню.

Номер можно набирать, диктуя цифры по одной или цепочками, состоящими не более чем из 9 цифр. Выбор нужной записи из голосовой телефонной книги можно сделать либо одной фразой, либо при помощи голосового меню. Существует также возможность голосового набора номера аварийной службы.

Комплекс DiVo может работать в таком режиме, когда он исполняет голосовые команды, произнесенные любым диктором (команды управления можно произносить на русском, английском и немецком языке). При этом предусмотрен авторизованный доступ к голосовой телефонной книге, ограничивающей использование комплекса посторонними людьми.

Что касается «пробуждения» комплекса DiVo, то для этого можно использовать голосовую команду или просто нажать соответствующую кнопку, расположенную на корпусе центрального блока.

Во время простоя комплекс автоматически выключается, что предотвращает ложные срабатывания на посторонние звуковые сигналы. Существует также возможность включения и использование некоторых функций кнопкой, смонтированной на колонке рулевого управления автомобиля.

Комплекс DiVo может работать в режиме полного дуплекса. При этом он обеспечивает подавление эха и высокоэффективную очистку сигнала от шумов. Это устраняет влияние на работу системы шумов, сопровождающих движение транспорта, гул колес и ветра, шум кондиционера или обогревателя и др.

Кроме того, на время телефонного разговора предусмотрено автоматическое отключение звука автомобильной аудиосистемы.

18-ЛЕКЦИЯ. МЕТОДЫ СИНТЕЗА РЕЧИ

План:

1. Методы синтеза речи.

В то время как задача распознавания речи очень сложна и решена лишь отчасти, задача синтеза речи намного проще (хотя и там есть немало проблем, ждущих своего решения).

В обиходе Вам наверняка приходилось сталкиваться с различными системами синтеза речи. Вот несколько примеров.

Служба 100 Московской городской телефонной сети, сообщающая по телефону текущее время, использует синтезатор речи. Диктор наговаривает цифры и такие слова, как «часов», «минут» и «секунд», а компьютер формирует речевое представление текущего времени.

Технологии синтеза речи применяются в метро при объявлении остановок.

Владельцы мобильных телефонов могут общаться с автоматической сервисной службой для определения остатка средств на счету, переключения тарифных планов, подключения или отключения услуг и пр. Сервисная служба общается голосом с применением технологий синтеза речи.

Выпущено немало детских игрушек, «говорящих» человеческим голосом. В этих игрушках также применяются простейшие синтезаторы речи или цифровые магнитофоны.

Синтезаторы речи применяются в различных голосовых системах предупреждения, устанавливаемых в автомобилях и самолетах. Такие системы позволяют привлечь внимание человека к возникновению той или иной критической ситуации, не отвлекая его от процесса управления автомобилем, самолетом или другим аналогичным средством.

Также разработано немало компьютерных программ, способных читать голосом содержимое текстовых файлов или текст, расположенный в окнах приложений. Эти системы могут оказаться полезными тем, у кого ослаблено или полностью отсутствует зрение.

Две модели синтеза речи

Все существующие в настоящее время методы синтеза человеческой речи основаны на использовании двух моделей — *модели компилятивного синтеза* и *формантно-голосовой модели* [3].

Рассмотрим вкратце особенности этих моделей.

Модель компилятивного синтеза

Модель компилятивного синтеза предполагает синтез речи путем конкатенации (составления) записанных образцов отдельных звуков, произнесенных диктором.

При использовании этой модели составляется база данных звуковых фрагментов, из которых в дальнейшем будет синтезироваться речь.

На первый взгляд этот подход не должен вызывать особых затруднений.

Действительно, пользуясь микрофоном и звуковым редактором, например, редактором GoldWave, описанным в 3 главе нашей книги, Вы можете создать набор файлов различных звуковых фрагментов, а затем сохранить их содержимое в базе данных.

Создавая звуковые WAV-файлы с текстовыми сообщениями, можно озвучить операционную систему Microsoft Windows и многие ее приложения, такие как почтовые программы, инструментальные средства разработки и пр.

Если Вы умеете создавать программы, то для Вас не составит труда сделать простейший синтезатор речи, составляющий фразы из отдельных фрагментов. Информацию, необходимую для составления программ, работающих со звуком, можно найти, например, в [10], а также в библиотеке MSDN (<http://msdn.microsoft.com>).

Модель компилятивного синтеза подходит, главным образом, только в простейших случаях, когда синтезатор должен произносить относительно небольшой и заранее известный набор фраз. При этом обеспечивается довольно высокое качество речи. Впрочем, этот факт не слишком удивителен, если вспомнить, что для синтеза используется естественная человеческая речь.

Тем не менее, на стыке составляемых звуковых фрагментов возможны интонационные искажения и разрывы, заметные на слух. Кроме того, создание крупной базы данных звуковых фрагментов, учитывающей все особенности произношения фонем и аллофонов с разными интонациями, представляет собой сложную и кропотливую работу.

Формантно-голосовая модель

Формантно-голосовая модель основана на моделировании речевого тракта человека, о котором мы подробно рассказывали в 1 главе нашей книги.

Эта модель может быть реализована с применением нейронных сетей и допускает самообучение. К сожалению, ввиду сложности точного моделирования особенностей речевого тракта, а также учета интонационной модуляции речи формантно-голосовая модель обладает относительно низкой точностью синтезируемых звуков речи. Тем не менее, современные программы синтеза речи, построенные с использованием этой модели, синтезируют вполне разборчивую речь и могут применяться в ряде случаев.

Заметим, что системы голосового предупреждения о возникновении аварийных ситуаций лучше строить с использованием модели компилятивного синтеза, так как разборчивость речи в таких системах выходит на передний план.

Что же касается «бытовых» синтезаторов речи, то в них можно с успехом применять и формантно-голосовую модель. Схематически эта модель показана на рис. 7-1 [3].

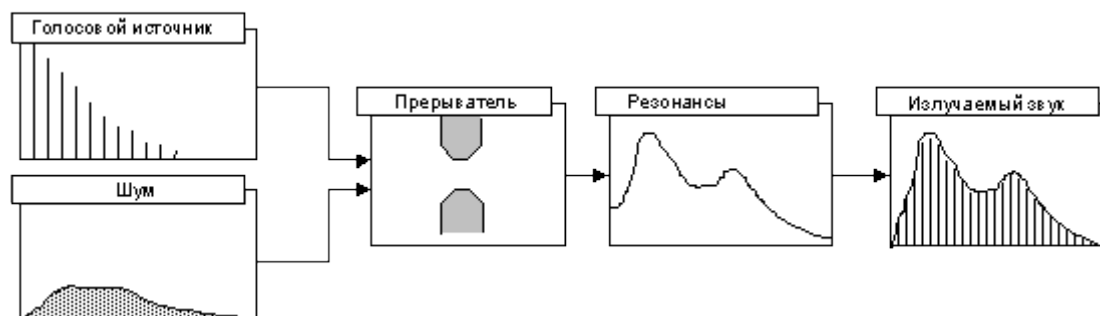


Рис. 7-1. Формантно-голосовая модель синтеза речи

При построении модели в [3] использовались данные об артикуляционном аппарате человека, а также данные фонетики и лингвистики. Как видите, в качестве исходного сигнала применяется комбинация голосового источника и генератора шума. Прерыватель и резонансное устройство моделирует работу

речевого тракта. В результате этого моделирования образуется излучаемый звук речи.

При этом для достижения компромисса между качеством модели и ее сложностью были выбраны следующие основные параметры исследуемой системы:

- ☐ частота основного тона;
- ☐ частота шума;
- ☐ количество формант;
- ☐ центральная частота каждой форманты;
- ☐ вклад каждой форманты.

Частота основного тона определяет высоту голоса. Этот параметр не должен вызывать у Вас никаких вопросов. Что же касается частоты шума, то здесь нужно сделать пояснение.

Как замечает автор работы [3], образование шума представляет собой достаточно сложный процесс, зависящий от многих факторов, таких как давление и скорость воздушной струи, геометрической формы воздушного тракта, акустических свойств материала и пр. Чтобы полностью смоделировать шум речи на физическом уровне, необходимо создать точную модель речевого аппарата человека, что представляет собой очень сложную задачу.

В качестве альтернативы автор работы [3] использует белый шум, спектр которого распределен по некоторому закону (например, по Гауссу) относительно некоторой центральной частоты. При этом закон распределения подбирается экспериментально, а частотой шума в этом случае является упомянутая выше центральная частота.

Количество активных формант, участвующих в образовании речи, выбирается в [3] экспериментально, причем в качестве ориентировочного значения используется 4.

Так как форманта представляет собой резонанс в речевом тракте, у неё есть *частота резонанса* и *оггибающая*. Вид оггибающей также определяется экспериментально, в первом приближении это Гауссово распределение.

Вклад каждой форманты определяет, насколько сильно форманта воздействует на основной сигнал.

Все приведенные выше параметры, кроме количества формант, изменяются в процессе образования речи для получения различных звуков. Хотя для более качественного синтеза речи необходимо строить более детальную модель, приведенные в [3] параметры достаточны для того, чтобы синтезированные звуки были разборчивы.

Синтез речи с помощью нейронной сети

В [3] описан процесс синтеза речи с использованием нейронной сети. Структура системы SAS, содержащей в своем составе эту сеть, была представлена в главе 5 на рис. 5-2.

Для исследования формантно-голосовой модели синтеза речи был создан инструмент Модель синтеза, в котором ручным заданием параметров можно синтезировать практически любой гласный или шипящий звук. Также в [3] приводятся уже готовые образцы некоторых звуков (в форме параметров модели).

Алгоритм синтеза речи

Процесс синтеза речи выглядит следующим образом. Уровни выходов нейронов эффекторного слоя нейросети при помощи карты эффекторов (рис. 5-2) преобразуются в значения выбранных параметров модели синтеза. Карта эффекторов определяет соответствие между каждым нейроном эффекторного слоя и конкретным параметром модели синтеза, а также предельные значения каждого параметра. Число эффекторов и число параметров модели может не совпадать. Если параметру не соответствует ни один эффектор, используется некоторое фиксированное значение (значение по умолчанию).

Далее по текущему состоянию модели синтезируется сигнал в пространстве частот: генерируется линейка частот, представляющих голосовой источник. На эту линейку частот накладывается формантная структура (резонансы). Для синтеза шума используется генератор случайной амплитуды и фазы.

На последнем этапе выполняется обратное преобразование Фурье для получения звуков речи.

При обучении системы формировались нейронные ансамбли для каждого звука из обучающей последовательности а, б, в, г, д. Затем проводилось обучение синтезу. В результате в эффекторном слое установились правильные связи с символьным слоем.

Система успешно обучилась синтезу — синтезируемые звуки в точности соответствуют тонам из обучающей выборки. На рис. 7-2 показана обучающая выборка, а на рис. 7-3 — результат синтеза.

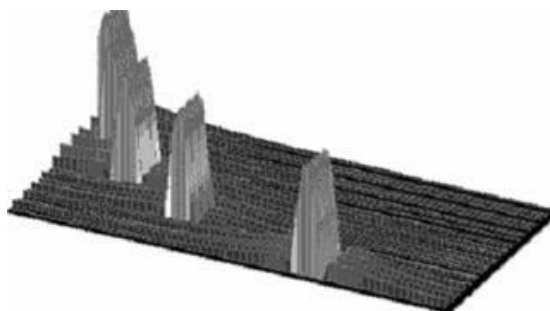


Рис. 7-2. Обучающая выборка

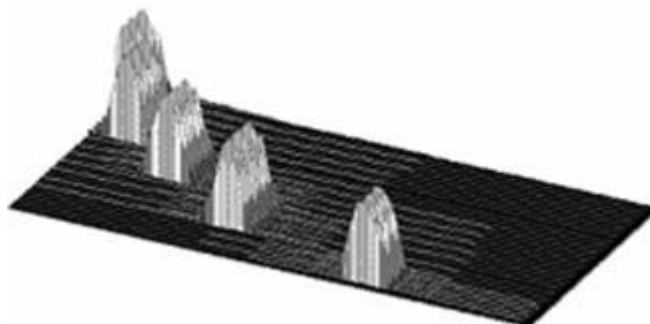


Рис. 7-3. Результат синтеза

На рис. 7-4 показан результат обучения синтезу звуков а,и,о,у.

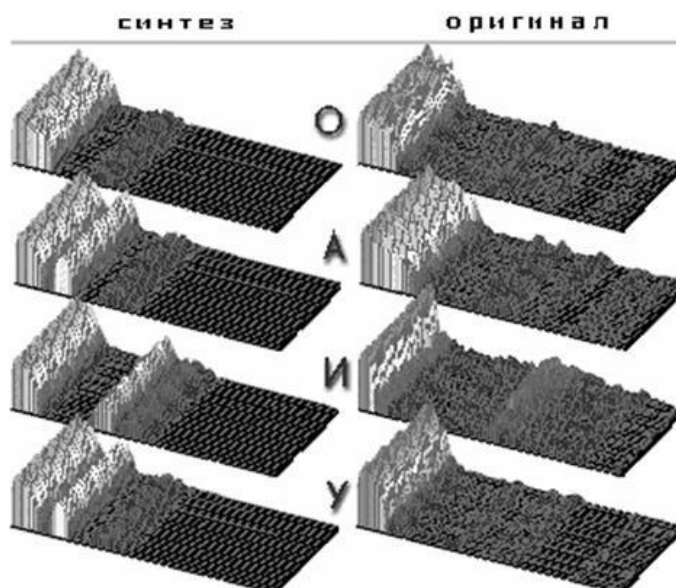


Рис. 7-4. Результат обучения синтезу

Как видите, спектрограммы синтезированных звуков близки к спектрограммам оригинальных звуков, хотя видны и отличия.

Более подробную информацию об условиях проведения экспериментов Вы найдете в [3].

Ограничения использованного алгоритма

Автор работы [3] отмечает, что в этом алгоритме узким местом является размер окна дискретного преобразования Фурье ДПФ. В данной модели синтезируются статичные звуки, при этом не происходит изменение параметров в процессе синтеза.

В реальной же речи параметры звука меняются при переходе от одного звука к другому, причем меняются непрерывно. Очевидно, при

использовании окон ДПФ такой результат получить невозможно — в пределах окна параметры звука меняться не будут. Теоретически благодаря полной обратимости дискретного преобразования Фурье возможно получить спектр для любого сигнала, в том числе и с динамически меняющимися параметрами.

Поэтому для генерации звука с изменяющимися параметрами нужно сокращать размер окна ДПФ или брать не весь сгенерированный кадр, а только его часть (не забывая при этом синхронизировать фазу сигнала). В идеале размер кадра можно свести к одному отсчету дискретизации по времени. Этот способ генерации речи дает лучшие результаты по сравнению с ДПФ, но работает гораздо медленнее ДПФ. В системе SAS можно выбрать используемый способ генерации.

1.2. Темы практических занятий.

1. Изучение звуков, частота звука, генерация звука на компьютере
2. Дискретизация звука. Хранение аудио сигналов в памяти компьютера
3. Обработка аудио сигналов в звуковых редакторах
4. Синтез речи на компьютере
5. Распознавание речи с помощью Google
6. Системы голосового управления

1.3. Темы лабораторных занятий.

1. Работа с генератором звука
2. Конвертация аудио форматов
3. Работа с редактором звука Wavosaur
4. Работа с синтезаторами речи
5. Работа с распознаванием речи системы Google
6. Управление компьютером с помощью голоса

1.3. Темы самостоятельных работ.

1. Спектральный анализ звуков.
2. Вейвлет-преобразования
3. Современные системы распознавания речи
4. Синтезаторы речи
5. Голос и речь человека
6. Форматы хранения аудио-сигналов
7. Нейронные сети в распознавании речи
8. Распознавание голоса по образцу
9. Системы голосового управления
10. Работа с аудио-редакторами

1.4. Глоссарий

“Архитектура” и “инженерия”, как виды человеческой деятельности, существовали задолго до появления компьютерных технологий. Прежде всего, эти виды деятельности связывал процесс создания проекта — прототипа, прообраза предполагаемого или возможного объекта.

Архитектура ПО — это артефакт, представляющий собой результат процесса разработки программного обеспечения.

В состав архитектурного проекта ПО входят: описание элементов, из которых состоит данная система, схемы взаимодействий между этими элементами, документация образцов (patterns), на основе которых осуществляется их компоновка, а также список и содержание ограничений (требований), характерных для этих образцов.

В качестве иллюстративных средств выражения характеристик ПО, в архитектурном проекте используются различные нотации – блок-схемы (схемы алгоритмов), ER-диаграммы, UML- диаграммы, DFD-диаграммы, а также макеты.

План реализации, составленный архитектором ПО, регламентирует способы получения конкретных систем из общей архитектуры и компонентов.

При отсутствии однозначной методики по применению ранее разработанного кода, кроме опыта и интуиции разработчиков, имеется наука такого программирования, являющаяся частью программной инженерии. Она называется «Порождающее программирование (Generative programming)».

Идеи порождающего программирования основываются на проектировании и построении порождающих моделей для семейств систем с целью генерирования по этим моделям конкретной системы.

Программная среда, которая осуществляет сборку готового к применению программного продукта, в том числе: программной системы, компонента, класса, процедуры и тому подобных частей системы на основе высокоуровневой спецификации, называется “генератор”.

Генераторы представляют собой множество различных технологий, включая препроцессоры, метафункции, компиляторы, генерирующие классы и процедуры, генераторы кода CASE-систем (инструментов автоматизированного проектирования и создания программ), трансформационные компоненты и многое другое.

Ярким и пока единственным примером реализаций таких систем является система IP, разработанная Чарльзом (Каролом) Симони (Charles (Karoly) Simonyi) в период его работы в корпорации Microsoft. Работа в среде метапрограммирования IP (Intentional Programming) иногда называется ментальное или интенциональное программирование.

Язык UML является формальным языком спецификаций и отличается тем самым от синтаксиса традиционных формально-логических языков и языков программирования.

Концепцией MDA является описание представления алгоритмов на языке моделирования с последующим автоматическим преобразованием моделей в компьютерный код, причем программирование на базе моделей предполагает, что проектировщики ПО прежде всего создают наиболее подходящую модель, не "привязываясь" к платформе, на которой система будет реализована.

Автоматизация архитектурного проектирования программного обеспечения основывается на применении инструментальных программных средств, которые принято называть CASE (Computer-Aided Software/System Engineering).

Формируемая платформенно-независимая модель создается на языке унифицированного моделирования UML.

После создания модели PIM, создаются одна или несколько платформенно-зависимых моделей, так называемые PSM (Platform Specific Model), назначение которых обеспечение интеграции PIM с одной или несколькими технологиями разработки программных продуктов.

Автоматизация архитектурного проектирования программного обеспечения основывается на применении инструментальных программных средств, которые принято называть CASE (Computer-Aided Software/System Engineering).

Библиотека STL состоит из пяти основных видов компонентов: Алгоритм (Algorithm), который определяет вычислительную процедуру. Контейнер (Container), назначение которого управлять набором объектов в памяти. Итератор (Iterator), который обеспечивает средство доступа к содержимому контейнера для алгоритма. Функциональный объект (Function object), который инкапсулирует функцию в объекте для её использования другими компонентами. Адаптер (Adaptor), который настраивает компонент для обеспечения различного интерфейса.

2.1. Учебная программа.

2.2. Рабочая программа.

**МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ УЗБЕКИСТАН**

**МИНИСТЕРСТВО ПО РАЗВИТИЮ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И
КОММУНИКАЦИЙ РЕСПУБЛИКИ УЗБЕКИСТАН**

**НУКУССКИЙ ФИЛИАЛ ТОШКЕНТСКОГО УНИВЕРСИТЕТА
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
ИМЕНИ МУХАММАДА АЛ-ХОРАЗМИЙ**

Зарегистрирован

№ _____

«УТВЕРЖДАЮ»

Зам. директор по учебной и
научной работе

_____ Х.Сейткамалов

«_____» _____ 2018 г.

РАБОЧАЯ ПРОГРАММА

дисциплины

“Алгоритмы распознавания речи”

Область знаний: 300 000 – Услуг техническая производства

Область образования: 330 000 – Компьютерные технологии и информатика

Направления образования: 5330300 – Компьютерный инжиниринг

НУКУС-2018

Рабочая программа составлена на основании типовой программы “Алгоритмы распознавания речи”.

Составитель:

Юлдашев К – ассистент кафедры “Программный инжиниринг” Нукусский филиал ТУИТ имени Мухаммада ал-Хорезмий.

Рецензия:

Б.Ш.Алламуратов – к.ф-м.н. кафедры Информационно-образовательные технологии” Нукусский филиал ТУИТ имени Мухаммада ал-Хорезмий.

Рабочая программа дисциплины «Алгоритмы распознавания речи» одобрена на собраниях кафедры (“___”от “___” – августа 2018 года), и передан на совет факультета.

Зав. кафедры: _____ **проф. Н.У.Утеулиев**

Рабочая программа дисциплины «Алгоритмы распознавания речи» утверждена на совете факультета “Компьютерный инжиниринг” (протокол “___” от “___” – августа 2018 г.).

Председатель совета факультета: _____ **д.т.н К.Сейтназаров**

Согласовано:

Начальник учебно-методического отдела: _____ **Ш.Ядгаров**

Введение.

Рабочая программа курса «Алгоритмы распознавания речи» описывает цели и задачи учебной дисциплины в соответствии с требованиями Государственного образовательного стандарта.

Цель изучения предмета и его задачи

Основные цели преподавания дисциплины «Алгоритмы распознавания речи» – является расширение и углубление знаний бакалавров в области программного обеспечения, разработки, построения и методы кодирования, техники и технологий.

Требования по предмету к знаниям и методикам студентов

В процессе обучения данной дисциплины на основе предъявляемых требований, бакалавры должны выполнить лабораторные занятия, готовить тексты по лекциям, обладать способностью к самостоятельному мышлению, использовать полученные навыки для написания курсовых работы и диссертационных работ.

Студенты должны уметь применять полученные знания на практических и лабораторных занятиях, а в дальнейшем для решения задач в различных отраслях производства.

После обучения предметы “Алгоритмы распознавание речи” студенты должны знать:

- Знать основные понятия обработки сигналов и понимать теоретическую базу компьютерного анализа речи.
- Уметь применять стандартные методы и модели к решению задач распознавания, синтеза и обработки речевых сигналов и получать обоснованные выводы.
- Иметь навыки (приобрести опыт) работы с основными принципами и методами обработки речевых данных, овладеть навыками применения речевых технологий и программ для анализа звука на ПЭВМ.

Взаимная связь с другими предметами в учебном плане.

Реализация программы основана на знаниях, приобретенных при обучении в бакалавриате по направлению «Программный инжиниринг» знаниям общих дисциплин «Программирование на C++», «Проектирование и моделирование компьютерных систем», «Принципы программирования», «Операционные системы и лаборатория», «Введение в программный инжиниринг», «Объектно-ориентированная программирования».

Новые педагогические технологии, используемые при изучении предмета

При освоении студентами курса «Алгоритмы распознавания речи» имеет особое значение применение новых информационно-педагогических технологий и современных методов обучения. Используются учебники, учебно-методические пособия, тексты лекций, раздаточные материалы, электронные материалы, лицензионные компьютерные программы, образцы информационных систем и технологий. Считается целесообразным проведение проблемных и открытых лекций, использование компьютерной техники, интерактивных методов проведения занятий и игр. При проведении лекций и лабораторных занятий применяются передовые педагогические технологии.

Ниже приведены факторы, связанные с процессом обучения и влияющие на качество образования: высокий научно-педагогический уровень преподавания; проведение проблемных лекций; организация урока в виде вопросов-ответов; использование новых педагогических и мультимедийных технологий; постановка перед студентами задач, вызывающих интерес и стремление к знаниям.

При планировании курса «Алгоритмы распознавания речи» были применены следующие концептуальные основы:

Человеко-ориентированное образование. Такое образование, в соответствии с содержанием, контролирует полноценное развитие всех участников образовательного процесса. При планировании образовательного процесса, он был ориентирован не только на единичных слушателей курса, имеется в виду подход на основе связанных с образовательными целями будущими специальностями.

Системный подход. Целостность и взаимосвязанность всех элементов образовательного процесса в технологии образования.

Подход, ориентированный на работоспособность. Улучшение качеств слушателей, активизация и интенсификация работоспособности слушателей и направление всех способностей, возможностей слушателей на процесс обучения.

Диалогический подход. Этот подход показывает необходимость появления учебной вовлеченности. В итоге у слушателей усиливается творческая активность и самовыражение.

Появление совместного образования. Означает необходимость совместной работы над оценкой достигнутых результатов со стороны слушателей и преподавателя, на основе принципов демократии и равенств.

Проблемное обучение. Один из методов активизации способностей слушателей методом проблемного представления содержания предмета. При этом обеспечивается понимание методов научного познания через объективное противоречие и его решение, возникновение диалектического доказательства и его развитие, их творческое применение на практике.

Использование современных средств представления информации – применение в процессе обучения компьютеров и различных информационных технологий.

Формы обучения: диалог, полилог, участие, фронтал основанный на совместной работе и самостоятельном обучении, коллектив и группа.

Инструменты обучения: традиционные формы обучения (лекции, тестовые лекции) а также компьютер и информационные технологии.

Инструменты коммуникации: прямое взаимодействие на основе оперативной обратной связи со студентами

Методы и инструменты обратной связи: наблюдение, блиц-вопрос, диагностика обучения на основе анализа рубежных, текущих и итоговых контрольных.

Методы и инструменты управления: планирование учебных занятий в виде технологической карты, отмечающей этапы занятия, совместные действия преподавателя и студента в достижении поставленных целей, занятия в аудитории, контроль за внеклассной работой.

Мониторинг и контроль: планирование наблюдения в процессе занятий и всего курса. В конце курса – проведение тестов или письменных работ и на их основе оценка знаний студентов.

**Распределения занятия дисциплины
«Алгоритмы распознавания речи» по лекционным и часовым:**

| № | Наименование темы | Лекция | Практическое занятия | Лабораторные занятия | Самостоятельная работа |
|---------------|--|---------------|---------------------------------|---------------------------------|-----------------------------------|
| 1 | Голосовое общение. Речь и слух человека. | 2 | 2 | | 2 |
| 2 | Звуковое оборудование компьютера. Преобразование звука в поток чисел | 2 | 2 | 2 | 2 |
| 3 | Анализ речевых сигналов. Осциллографические исследования | 2 | 2 | 2 | 4 |
| 4 | Введение в нейронные сети. Структуры нейронных сетей | 4 | | 2 | 2 |
| 5 | Методы распознавания речи. Распознавание по образцу | 4 | 2 | | 2 |
| 6 | Выделение лексических элементов речи | 4 | | 2 | 4 |
| 7 | Предварительная обработка звуковых сигналов | 4 | 2 | | 2 |
| 8 | Выделение информативных признаков речевого сигнала | 2 | | 2 | 4 |
| 9 | Выделение фонем и аллофонов | 2 | 2 | | 2 |
| 10 | Уровни распознавания слитной речи | 2 | | 2 | 2 |
| 11 | Применение нейронных сетей для распознавания речи | 2 | 2 | 2 | 4 |
| 12 | Применение вейвлет-преобразований при распознавании речи | 2 | 2 | 2 | 4 |
| 13 | Системы голосового управления | 2 | 2 | | 2 |
| 14 | Методы синтеза речи | 2 | | 2 | 4 |
| ВСЕГО: | | 36 | 18 | 18 | 40 |

Основная часть:

Содержание теоретических занятий предмета.

В основной части (лекция) приводятся темы предмета в логической последовательности. Суть каждой темы раскрывается с помощью основных понятий и тезисов.

По теме студенты должны получить необходимые знания и навыки согласно государственному стандарту.

В основной части рекомендуется раскрывать актуальность темы, ее соответствие требованиям работодателей и производителей, соответствие социально-политическим и демократическим направлениям.

Лекционные занятия

1. Голосовое общение. Речь и слух человека.

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; О5; Д2; Д3; Д4.

2. Звуковое оборудование компьютера. Преобразование звука в поток чисел

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; Д2.

3. Анализ речевых сигналов. Осциллографические исследования

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; Д2; Д3.

4. Введение в нейронные сети. Структуры нейронных сетей

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; Д1; Д2; Д3.

5. Методы распознавания речи. Распознавание по образцу

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Б/Б/Б жадвали, мунозара, Венн диаграммаси, Т-схема, ўз-ўзини назорат

Литературы: О3; О4; О5; Д2; Д3; Д4.

6. Выделение лексических элементов речи

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Ажурали арра, бумеранг, усули, мунозара, ўз-ўзини назорат.

Литературы: О3; О4; О5; Д2; Д3; Д4.

7. Предварительная обработка звуковых сигналов

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Блитс, усули, мунозара, ўз-ўзини назорат.

Литературы: О3; О4; О5; Д1; Д2; Д3; Д4.

8. Выделение информативных признаков речевого сигнала

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; Д2; Д3.

9. Выделение фонем и аллофонов

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Маъруза, намойиш этиш, блиц-сўров, гуруҳларда ишлаш методи.

Литературы: О3; О4; О5; Д2; Д3; Д4.

10. Уровни распознавания слитной речи

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Маъруза, намойиш этиш, “Блиц-сўров” методлари.

Литературы: О4; О5; Д2; Д3; Д4.

11. Применение нейронных сетей для распознавания речи

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим. Маъруза, намойиш этиш, “Блиц-сўров” методлари.*

Литературы: О3; О4; Д2; Д3.

12. Применение вейвлет-преобразований при распознавании речи

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; О5; Д2; Д3; Д4.

13. Системы голосового управления

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; О5; Д2; Д3; Д4.

14. Методы синтеза речи

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О3; О4; О5; Д2; Д3; Д4.

Календарно-тематический план лекционных занятий по предмету «Алгоритмы распознавания речи»

| П/п | Темы лекционных занятий | Кол. часов |
|--------|--|------------|
| 1 | Голосовое общение. Речь и слух человека. | 2 |
| 2 | Звуковое оборудование компьютера. Преобразование звука в поток чисел | 2 |
| 3 | Анализ речевых сигналов. Осциллографические исследования | 2 |
| 4 | Введение в нейронные сети. Структуры нейронных сетей | 2 |
| 5 | Введение в нейронные сети. Структуры нейронных сетей | 2 |
| 6 | Методы распознавания речи. Распознавание по образцу | 2 |
| 7 | Методы распознавания речи. Распознавание по образцу | 2 |
| 8 | Выделение лексических элементов речи | 2 |
| 9 | Выделение лексических элементов речи | 2 |
| 10 | Предварительная обработка звуковых сигналов | 2 |
| 11 | Предварительная обработка звуковых сигналов | 2 |
| 12 | Выделение информативных признаков речевого сигнала | 2 |
| 13 | Выделение фонем и аллофонов | 2 |
| 14 | Уровни распознавания слитной речи | 2 |
| 15 | Применение нейронных сетей для распознавания речи | 2 |
| 16 | Применение вейвлет-преобразований при распознавании речи | 2 |
| 17 | Системы голосового управления | 2 |
| 18 | Методы синтеза речи | 2 |
| ВСЕГО: | | 36 |

Рекомендуемые темы практических занятий.

На практических занятиях студенты изучают общее устройство компьютера, его основные и дополнительные устройства и принципы их работы.

В процессе обучения студенты получают знания об устройстве современных компьютерных архитектур, о современных процессорах и архитектурах шин, о центральной процессоре и последовательности процессорных команд, архитектуре ввода-вывода, шинах ввода-вывода, об устройстве видеопамяти и телекоммуникации, архитектурах параллельных компьютеров, о структуре мультипроцессоров и мультикомпьютеров, о разных этапах построения компьютерной архитектуры, этапы микроархитектуры, архитектура набора команд, а также основные понятия о прикладной системе и языке ассемблер.

При подготовке практических заданий со стороны преподавателей кафедры вносятся предложения. В том числе, рекомендуется на основе учебников и учебных пособий повышать знания студентов и широко использовать раздаточные материалы.

Применяющие образовательные технологии: *мозговой штурм, групповое обучение.*

Литературы: О3; О4; О5; Д1; Д2; Д3.

1. Изучение звуков, частота звука, генерация звука на компьютере

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О1; О2; Д2; Д3.

2. Дискретизация звука. Хранение аудио сигналов в памяти компьютера

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О2; О3; Д2; Д3.

3. Обработка аудио сигналов в звуковых редакторах

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О2; О3; Д2; Д3.

4. Синтез речи на компьютере

Применяющие образовательные технологии: *БББ, Инсерт, ўз-ўзини назорат.*

Литературы: О1; О2; О3; Д2; Д3.

5. Распознавание речи с помощью Google

Применяющие образовательные технологии: *Маъруза, намойиш этиш, блиц-сўров.*

Литературы: О2; О3; О4; Д1; Д2; Д3; Д4.

6. Системы голосового управления

Применяющие образовательные технологии: *диалогик ёндошув, муаммоли таълим.*

Литературы: О2; О3; О4; Д1; Д2; Д3; Д4;

Календарно-тематический план практических занятий по предмету «Алгоритмы распознавания речи»

| № | Наименование темы практических занятий | Кол. часов |
|----------|--|-------------------|
| 1. | Изучение звуков, частота звука, генерация звука на компьютере | 4 |
| 2. | Дискретизация звука. Хранение аудио сигналов в памяти компьютера | 4 |
| 3. | Обработка аудио сигналов в звуковых редакторах | 4 |
| 4. | Синтез речи на компьютере | 2 |
| 5. | Распознавание речи с помощью Google | 2 |

| | | |
|---------------|-------------------------------|-----------|
| 6. | Системы голосового управления | 2 |
| ВСЕГО: | | 18 |

**Календарно-тематический план лабораторных занятий по предмету
«Алгоритмы распознавания речи»**

| № | Наименование темы лабораторных занятий | Кол. часов |
|---------------|---|-------------------|
| 1 | Работа с генератором звука | 4 |
| 2 | Конвертация аудио форматов | 4 |
| 3 | Работа с редактором звука Wavosaur | 4 |
| 4 | Работа с синтезаторами речи | 2 |
| 5 | Работа с распознаванием речи системы Google | 2 |
| 6 | Управление компьютером с помощью голоса | 2 |
| ВСЕГО: | | 18 |

Форма и содержания организации самостоятельных работ.

Основная цель выполнения самостоятельных работ – организация непрерывного процесса обучения студентов под прямым руководством преподавателя, укрепление полученных знаний и навыков, подготовка к следующим занятиям, организация культуры умственного труда и самостоятельного получения новых знаний.

Самостоятельные работы (5 семестр, 48 часов)

| № | Тема самостоятельной работы | Задание | Срок | Объем (в часах) |
|----------|--|-----------------------------------|-------------|------------------------|
| 1. | Спектральный анализ звуков | Выполнение самостоятельной работы | 1-неделя | 4 |
| 2. | Вейвлет-преобразования | Выполнение индивидуальных заданий | 2-неделя | 4 |
| 3. | Современные системы распознавания речи | Выполнение самостоятельной работы | 2,3-недели | 4 |
| 4. | Синтезаторы речи | Выполнение самостоятельной работы | 3,4-недели | 4 |
| 5. | Голос и речь человека | Выполнение самостоятельной работы | 5,6-неделя | 4 |

| | | | | |
|--------------|-------------------------------------|-----------------------------------|------------|-----------|
| 6. | Форматы хранения аудио-сигналов | Выполнение индивидуальных заданий | 6-неделя | 4 |
| 7. | Нейронные сети в распознавании речи | Выполнение самостоятельной работы | 6,7-неделя | 4 |
| 8. | Распознавание голоса по образцу | Выполнение индивидуальных заданий | 7-неделя | 4 |
| 9. | Системы голосового управления | Выполнение самостоятельной работы | 8,9-недели | 4 |
| 10. | Работа с аудио-редакторами | Выполнение самостоятельной работы | 10-неделя | 4 |
| Итого | | | | 40 |

Информационно-методическое обеспечение

В процессе преподавания дисциплины применялись современные методы образования, педагогические и информационно-коммуникационные технологии:

- на всех лекциях применяются презентации с помощью современных компьютерных и электронно-дидактических технологий;

- во время практических занятий широко применяются новейшие компьютерные устройства;

- во время практических занятий широко применяются мультимедийные технологии, во время всех практических занятий применяются педагогические технологии «мозговой штурм», «групповое рассуждение».

Критерии оценки знаний студентов на основе рейтинговой системы по предмету «Алгоритмы распознавания речи»

Во время первого занятия по предмету студентам объявляется информация по рейтинговой таблице предмета, способам и формам оценки, количество и максимальный балл каждой контрольной, в том числе по баллам текущих и рубежных контрольных.

Для соответствия знаний и успеваемости студентов Государственным стандартам проводятся следующие виды контроля:

Текущий контроль (ТК) – способ определения и оценки уровня знаний студента по темам предмета и практическим навыкам. Текущий контроль может проводиться, исходя из качеств изучаемого предмета, в виде устного опроса, тестирования, диспута, контрольной работы, коллоквиума, проверки домашних заданий, а также в других подобных формах.

Рубежный контроль (РК) – способ определения и оценки уровня теоретических знаний и практических навыков студента, проводимый после прохождения соответствующего раздела (включающего в себя несколько тем) учебной программы. Рубежный контроль проводится дважды за семестр, а его форма (письменно, устно, тестирование, и т.д.) определяется исходя из объема выделенных часов на данный предмет.

Итоговый контроль (ИК) – способ оценки уровня усвоения студентами теоретических знаний и практических навыков по точным предметам по итогам семестра. Итоговый контроль проводится в форме письменной работы, основанной на опорных понятиях.

Итоговый контроль проводится при непосредственном участии комиссии, составленной заведующим кафедрой. При нарушении порядка проведения итогового контроля, результаты итогового контроля могут быть признаны недействительными. В этом случае итоговый контроль проводится заново.

Итоговый контроль проводится под руководством комиссии, составленной на основании приказа руководителя высшего учебного заведения. При нарушении порядка проведения итогового контроля, результаты итогового контроля могут быть признаны недействительными. В этом случае итоговый контроль проводится заново.

Рейтинговая система оценки знаний и навыков студентов основывается на успеваемости студентов, выраженной в баллах.

Показатель успеваемости студентов по курсу «Алгоритмы распознавания речи» за семестр оценивается по 100-балльной системе.

Эти 100 баллов распределяются по способам оценки следующим образом:

ИК – 30 баллов, остальные 70 баллов: **ТК** – 36 баллов, **РК** – 34 балла.

| Балл | Оценка | Уровень знаний студентов |
|--------|---------|--|
| 86-100 | Отлично | Вывод и принятие решение. Творческое мышление. Самостоятельно делать вывод. Умение пользоваться знаниями на практике. Объяснение сути. Знать и объяснить. Иметь представление. |
| 71-85 | Хорошо | Самостоятельно делать вывод. Умение пользоваться знаниями на практике. Объяснение сути. Знать и объяснить. Иметь представление. |
| 55-70 | Удов. | Объяснение сути. Знать и объяснить. Иметь представление. |
| 0-54 | Неудов. | Незнание предмета. |

Проходной балл по предмету составляет 55 баллов. Успеваемость ниже проходного балла не записывается в рейтинговую книжку.

Самостоятельные работы студентов по изучаемому предмету оцениваются в процессе текущих, рубежных и итоговых контрольных исходя из выделенных часов, в соответствии с выполнением порученных заданий.

Рейтинг студента по предмету определяется по следующей формуле:

$$R = \frac{V \cdot O'}{100}$$

где V – общий объем нагрузки, выделенной за семестр на данный предмет (в часах);

O' - успеваемость по предмету (в баллах).

Проходной балл по текущим и рубежным контрольным составляет 55 баллов. Студенты, набравшие баллы меньше проходного балла, не допускаются к итоговым контрольным.

Студенты, набравшие во время текущих контрольных (ТК) и рубежных контрольных (РК) больше 55 баллов, считаются успевающими по предмету и в дальнейшем допускаются до итогового контроля.

Общий балл, набранный студентом за семестр, в соответствии с правилами, высчитывается как сумма каждой контрольной.

Рубежный и итоговый контроли, в соответствии с календарным тематическим планом, проводится на основании рейтинговой контрольной таблицы, разработанной деканатом. Итоговый контроль проводится в последние 2 недели семестра.

Студентам, во время ТК и РК набравших баллы ниже проходных, а также студентам, которые по уважительным причинам не приняли участие в контрольных работах, дается срок до следующей ТК либо РК для повторной сдачи контрольной.

Студенты, в течение семестра набравшие менее 55 баллов по ТК и РК, либо в течение семестра набравшие по РК и ИК менее 55 баллов, считаются академическими должниками.

В случае несогласия с результатами контрольных, студент в течение одного дня после объявления результатов контрольной по предмету может обратиться в деканат с заявлением. В данном случае, на основании предложения декана факультета, по приказу ректора организовывается апелляционная комиссия, состоящая не менее чем из 3 (трех) человек.

Апелляционная комиссия, рассмотрев заявления студентов, в этот же день выносит решение.

Своевременное проведение контрольных работ согласно установленным правилам и их документирование должно проходить под контролем декана факультета, заведующего кафедрой, учебно-методического отдела и отдела внутреннего контроля и мониторинга.

Образцы критериев начисления баллов РК

| № | Показатель | Баллы РК | | |
|---------------------------|--|-----------|-------------|-------------|
| | | Макс | 1-РК | 2-РК |
| 1. | Посещаемость занятий. Активность на лекционных занятиях, наличие конспектов и их полнота. | 4 | 0-2 | 0-2 |
| 2. | Ответы на тестовые или письменные вопросы | 22 | 0-10 | 0-12 |
| 3. | Своевременное и качественное выполнение самостоятельных работ. Выполнение домашних работ по теме, успеваемость | 8 | 0-4 | 0-4 |
| Итого баллов по РК | | 34 | 0-16 | 0-18 |

Образцы критериев начисления баллов ТК

| № | Показатель | Баллы ТК |
|---|------------|----------|
|---|------------|----------|

| | | макс | 1-ТК | 2-ТК |
|--------------------|---|-----------|-------------|-------------|
| 1. | Посещаемость и успеваемость по занятиям. Активность на лабораторных занятиях, наличие тетради для лабораторных работ и их состояние | 4 | 0-2 | 0-2 |
| 2. | Своевременное и качественное выполнение лабораторных работ | 24 | 0-12 | 0-12 |
| 3. | Своевременное и качественное выполнение самостоятельных работ. Выполнение домашних работ по теме, успеваемость | 8 | 0-4 | 0-4 |
| Итого по ТК | | 36 | 0-18 | 0-18 |

В случае, если итоговый контроль проводится в виде письменной работы, итоговый контроль проводится в виде 30-балльной письменной работы с вариантами.

В случае, если итоговый контроль по предмету проводится в виде письменной работы, итоговый контроль проводится согласно приведенной ниже таблице:

| № | Показатель | Баллы ИК | |
|--------------|---|--------------|--------------------|
| | | Максимальный | Диапазон изменения |
| 1. | Письменная контрольная работа по предмету | 30 | 0-30 |
| Итого | | 30 | 0-30 |

Критерии оценки письменных работ на итоговых контрольных

В случае, если итоговая контрольная работа проводится в виде письменной работы, проверка проводится методом множества вариантов. Каждый вариант состоит из 3 теоретических и 2 практических заданий. Теоретические вопросы включают в себя ключевые слова и словосочетания по предмету, и охватывают собой все темы предмета.

На каждый теоретический вопрос дается оценка по 6-балльной шкале (от 0 до 6 баллов). Практические задания оцениваются от 0 до 6 баллов. Студент может набрать максимум 30 баллов.

Для определения показателя успеваемости по письменной работе, набранные баллы по каждому ответу на поставленный в варианте вопрос складываются, и сумма считается оценкой за итоговую контрольную работу.

СПИСОК РЕКОМЕНДОВАННЫХ ЛИТЕРАТУР

Основная литература и учебные пособия:

1. Фролов А., Фролов Г. Синтез и распознавание речи. Современные решения
2. Frederick Jelinek. Statistical Methods for Speech Recognition. MIT Press, Cambridge, MA, 2008
3. John N. Holmes, Wendy J. Holmes, "Speech Synthesis and Recognition", Taylor & Francis (2001), 2nd edition

Дополнительные литературы:

1. Савченко, В.В. Автоматическое распознавание изолированных слов методом обеляющего фильтра
2. Савченко, В.В. Фонетический анализ речи методом переменного дерева
3. Digital Speech Processing, Synthesis, and Recognition by Sadaoki Furui published in 2012 by Tokai University press.
4. Trends in Speech recognition by Wayne A. Lea. 1980 by Prentice Hall..

Интернет сайты:

1. <http://www.frolov-lib.ru>
2. <http://www.cs.nyu.edu/~eugenew/asr13/>
3. <http://www.ece.ucsb.edu/Faculty/Rabiner/ece259/speech%20recognition%20course.html>
4. <http://www.elsevier.com/books/readings-in-speech-recognition/waibel/978-1-55860-124-6>

2.4. Тесты и варианты для контроля знания студентов.

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 1

1. Речь и слух человека
2. Дискретное преобразование Фурье
3. Предварительная обработка звуковых сигналов
4. Методы и программы синтеза речи
5. Применение вейвлет-преобразований

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 2

1. Речевой тракт человека
2. Логарифмическое сжатие спектра
3. Применение нейронных сетей для распознавания речи
4. Синтез речи с помощью нейронной сети
5. Две модели синтеза речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 3

1. Преобразование звука в поток чисел
2. Исследование спектра речевых звуков
3. Системы распознавания и обработки речи
4. Уровни распознавания слитной речи
5. Синтез речи с помощью нейронной сети

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 4

1. Аналогово-цифровой преобразователь
2. Понятие о нейронных сетях
3. Системы голосового управления
4. Выделение фонем и аллофонов
5. Программные реализации синтезаторов речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 5

1. Преобразование потока чисел в звук
2. Структуры нейронных сетей

3. Голосовое разграничение доступа
4. Уровни распознавания слитной речи
5. Синтезатор речи Govorka

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 6

1. Анализ речевых сигналов
2. Методы распознавания речи
3. Предварительная обработка звуковых сигналов
4. Методы и программы синтеза речи
5. Профессиональные системы анализа речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 7

1. Спектральное представление речи
2. Программы для диктовки текста
3. Применение нейронных сетей для распознавания речи
4. Синтез речи с помощью нейронной сети
5. Применение вейвлет-преобразований

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 8

1. Распознавание речи по образцу
2. Дискретное преобразование Фурье
3. Системы распознавания и обработки речи
4. Уровни распознавания слитной речи
5. Две модели синтеза речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 9

1. Речь и слух человека
2. Логарифмическое сжатие спектра
3. Системы голосового управления
4. Выделение фонем и аллофонов
5. Синтез речи с помощью нейронной сети

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 10

1. Речевой тракт человека
2. Исследование спектра речевых звуков
3. Голосовое разграничение доступа
4. Уровни распознавания слитной речи
5. Программные реализации синтезаторов речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 11

1. Преобразование звука в поток чисел
2. Понятие о нейронных сетях
3. Предварительная обработка звуковых сигналов
4. Методы и программы синтеза речи
5. Синтезатор речи Govorilka

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 12

1. Аналогово-цифровой преобразователь
2. Структуры нейронных сетей
3. Применение нейронных сетей для распознавания речи
4. Синтез речи с помощью нейронной сети
5. Профессиональные системы анализа речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 13

1. Преобразование потока чисел в звук
2. Методы распознавания речи
3. Системы распознавания и обработки речи
4. Уровни распознавания слитной речи
5. Применение вейвлет-преобразований

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 14

1. Анализ речевых сигналов
2. Программы для диктовки текста
3. Системы голосового управления
4. Выделение фонем и аллофонов
5. Две модели синтеза речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 15

1. Спектральное представление речи
2. Дискретное преобразование Фурье
3. Голосовое разграничение доступа
4. Уровни распознавания слитной речи
5. Синтез речи с помощью нейронной сети

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 16

1. Распознавание речи по образцу
2. Логарифмическое сжатие спектра
3. Предварительная обработка звуковых сигналов
4. Методы и программы синтеза речи
5. Программные реализации синтезаторов речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 17

1. Речь и слух человека
2. Исследование спектра речевых звуков
3. Применение нейронных сетей для распознавания речи
4. Синтез речи с помощью нейронной сети
5. Синтезатор речи Govorka

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 18

1. Речевой тракт человека
2. Понятие о нейронных сетях
3. Системы распознавания и обработки речи
4. Уровни распознавания слитной речи
5. Профессиональные системы анализа речи

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 19

1. Преобразование звука в поток чисел
2. Структуры нейронных сетей
3. Системы голосового управления

4. Выделение фонем и аллофонов
5. Применение вейвлет-преобразований

Итоговая контрольная по предмету «Алгоритмы распознавания речи» Заведующий кафедрой
Кафедра «Программный инжиниринг» _____ Н.Утеулиев

Вариант 20

1. Аналогово-цифровой преобразователь
2. Методы распознавания речи
3. Голосовое разграничение доступа
4. Уровни распознавания слитной речи
5. Две модели синтеза речи

2.5. Методические указания по применению критериев оценок.

Критерии оценки знаний студентов на основе рейтинговой системы по предмету «Алгоритмы распознавания речи»

Во время первого занятия по предмету студентам объявляется информация по рейтинговой таблице предмета, способам и формам оценки, количество и максимальный балл каждой контрольной, в том числе по баллам текущих и рубежных контрольных.

Для соответствия знаний и успеваемости студентов Государственным стандартам проводятся следующие виды контроля:

Текущий контроль (ТК) – способ определения и оценки уровня знаний студента по темам предмета и практическим навыкам. Текущий контроль может проводиться, исходя из качеств изучаемого предмета, в виде устного опроса, тестирования, диспута, контрольной работы, коллоквиума, проверки домашних заданий, а также в других подобных формах.

Рубежный контроль (РК) – способ определения и оценки уровня теоретических знаний и практических навыков студента, проводимый после прохождения соответствующего раздела (включающего в себя несколько тем) учебной программы. Рубежный контроль проводится дважды за семестр, а его форма (письменно, устно, тестирование, и т.д.) определяется исходя из объема выделенных часов на данный предмет.

Итоговый контроль (ИК) – способ оценки уровня усвоения студентами теоретических знаний и практических навыков по точным предметам по итогам семестра. Итоговый контроль проводится в форме письменной работы, основанной на опорных понятиях.

Итоговый контроль проводится при непосредственном участии комиссии, составленной заведующим кафедрой. При нарушении порядка проведения итогового контроля, результаты итогового контроля могут быть признаны недействительными. В этом случае итоговый контроль проводится заново.

Итоговый контроль проводится под руководством комиссии, составленной на основании приказа руководителя высшего учебного заведения. При нарушении порядка проведения итогового контроля, результаты итогового контроля могут быть признаны недействительными. В этом случае итоговый контроль проводится заново.

Рейтинговая система оценки знаний и навыков студентов основывается на успеваемости студентов, выраженной в баллах.

Показатель успеваемости студентов по курсу «Архитектура программного обеспечения. Построение и кодирование программного обеспечения» за семестр оценивается по 100-балльной системе.

Эти 100 баллов распределяются по способам оценки следующим образом:

ИК – 30 баллов, остальные 70 баллов: **ТК** – 36 баллов, **РК** – 34 балла.

| Балл | Оценка | Уровень знаний студентов |
|--------|---------|--|
| 86-100 | Отлично | Вывод и принятие решение. Творческое мышление. Самостоятельно делать вывод. Умение пользоваться знаниями на практике. Объяснение сути. Знать и объяснить. Иметь представление. |
| 71-85 | Хорошо | Самостоятельно делать вывод. Умение пользоваться знаниями на практике. Объяснение сути. Знать и объяснить. Иметь представление. |
| 55-70 | Удов. | Объяснение сути. Знать и объяснить. Иметь представление. |
| 0-54 | Неудов. | Незнание предмета. |

Проходной балл по предмету составляет 55 баллов. Успеваемость ниже проходного балла не записывается в рейтинговую книжку.

Самостоятельные работы студентов по изучаемому предмету оцениваются в процессе текущих, рубежных и итоговых контрольных исходя из выделенных часов, в соответствии с выполнением порученных заданий.

Рейтинг студента по предмету определяется по следующей формуле:

$$R = \frac{V \cdot O'}{100}$$

где V – общий объем нагрузки, выделенной за семестр на данный предмет (в часах);

O' - успеваемость по предмету (в баллах).

Проходной балл по текущим и рубежным контрольным составляет 55 баллов. Студенты, набравшие баллы меньше проходного балла, не допускаются к итоговым контрольным.

Студенты, набравшие во время текущих контрольных (ТК) и рубежных контрольных (РК) больше 55 баллов, считаются успевающими по предмету и в дальнейшем допускаются до итогового контроля.

Общий балл, набранный студентом за семестр, в соответствии с правилами, высчитывается как сумма каждой контрольной.

Рубежный и итоговый контроли, в соответствии с календарным тематическим планом, проводится на основании рейтинговой контрольной таблицы, разработанной деканатом. Итоговый контроль проводится в последние 2 недели семестра.

Студентам, во время ТК и РК набравших баллы ниже проходных, а также студентам, которые по уважительным причинам не приняли участие в контрольных работах, дается срок до следующей ТК либо РК для повторной сдачи контрольной.

Студенты, в течение семестра набравшие менее 55 баллов по ТК и РК, либо в течение семестра набравшие по РК и ИК менее 55 баллов, считаются академическими должниками.

В случае несогласия с результатами контрольных, студент в течение одного дня после объявления результатов контрольной по предмету может обратиться в деканат с заявлением. В данном случае, на основании предложения декана факультета, по приказу ректора организовывается апелляционная комиссия, состоящая не менее чем из 3 (трех) человек.

Апелляционная комиссия, рассмотрев заявления студентов, в этот же день выносит решение.

Своевременное проведение контрольных работ согласно установленным правилам и их документирование должно проходить под контролем декана факультета, заведующего кафедрой, учебно-методического отдела и отдела внутреннего контроля и мониторинга.

Образцы критериев начисления баллов РК

| № | Показатель | Баллы РК | | |
|---------------------------|--|-----------------|-------------|-------------|
| | | Макс | 1-РК | 2-РК |
| 1. | Посещаемость занятий. Активность на лекционных занятиях, наличие конспектов и их полнота. | 4 | 0-2 | 0-2 |
| 2. | Ответы на тестовые или письменные вопросы | 22 | 0-10 | 0-12 |
| 3. | Своевременное и качественное выполнение самостоятельных работ. Выполнение домашних работ по теме, успеваемость | 8 | 0-4 | 0-4 |
| Итого баллов по РК | | 34 | 0-16 | 0-18 |

Образцы критериев начисления баллов ТК

| № | Показатель | Баллы ТК | | |
|--------------------|---|-----------------|-------------|-------------|
| | | макс | 1-ТК | 2-ТК |
| 1. | Посещаемость и успеваемость по занятиям. Активность на лабораторных занятиях, наличие тетради для лабораторных работ и их состояние | 4 | 0-2 | 0-2 |
| 2. | Своевременное и качественное выполнение лабораторных работ | 24 | 0-12 | 0-12 |
| 3. | Своевременное и качественное выполнение самостоятельных работ. Выполнение домашних работ по теме, успеваемость | 8 | 0-4 | 0-4 |
| Итого по ТК | | 36 | 0-18 | 0-18 |

В случае, если итоговый контроль проводится в виде письменной работы, итоговый контроль проводится в виде 30-балльной письменной работы с вариантами.

В случае, если итоговый контроль по предмету проводится в виде письменной работы, итоговый контроль проводится согласно приведенной ниже таблице:

| № | Показатель | Баллы ИК | |
|-------|---|--------------|--------------------|
| | | Максимальный | Диапазон изменения |
| 1. | Письменная контрольная работа по предмету | 30 | 0-30 |
| Итого | | 30 | 0-30 |

Критерии оценки письменных работ на итоговых контрольных

В случае, если итоговая контрольная работа проводится в виде письменной работы, проверка проводится методом множества вариантов. Каждый вариант состоит из 3 теоретических и 2 практических заданий. Теоретические вопросы включают в себя ключевые слова и словосочетания по предмету, и охватывают собой все темы предмета.

На каждый теоретический вопрос дается оценка по 6-балльной шкале (от 0 до 6 баллов). Практические задания оцениваются от 0 до 6 баллов. Студент может набрать максимум 30 баллов.

Для определения показателя успеваемости по письменной работе, набранные баллы по каждому ответу на поставленный в варианте вопрос складываются, и сумма считается оценкой за итоговую контрольную работу.

