

Лабораторная работа

«Изучение фонетических, просодических и компиляционных закономерностей речевого сигнала»

Цель работы – выявить и изучить фонетические, просодические и компиляционные закономерности речевого сигнала.

Основные **задачи** работы следующие:

1. Закрепление основных понятий и характеристик речевого сигнала.
2. Приобретение навыков работы с программой *Sonic Foundry Sound Forge* или программы с подобными функциональными возможностями для изучения речевого сигнала.
3. Практическое исследование фонетических, просодических и компиляционных закономерностей речевого сигнала.

В отчет по работе необходимо включить:

- ✓ Результаты выполнения заданий части 1-3.
- ✓ Описание используемого программного обеспечения для обработки речевого сигнала.
- ✓ Выводы.

II. Описание речевого сигнала

Речь – это сложное физическое явление. Она образуется в результате прохождения воздушных потоков через речевой тракт человека. С акустической точки зрения **звук** – это колебательные движения среды. Сложность речи заключается в том, что это не один колебательный процесс. Речевой сигнал состоит из множества **гармонических составляющих**, т.е. колебаний, которые периодически изменяются во времени по синусоидальным законам (рис. 21) и описываются следующим уравнением [1] (*Лобанов Б.М..2006кн-Речев_И_ИС*):

$$S(t) = A \cdot \cos(F \cdot t + \phi)$$

, где A – амплитуда колебаний; F – частота ($F = 1 / T$); t – время; ϕ – фаза.

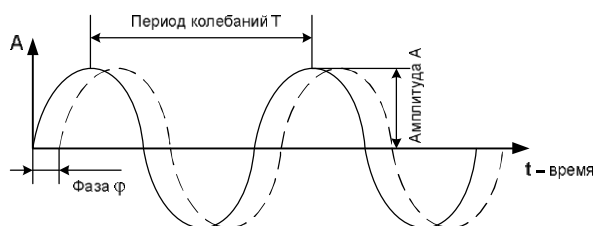


Рисунок 2 1 . Синусоидальное гармоническое колебание

Речевой сигнал, как и любой сложный сигнал, является результатом сложения нескольких гармонических составляющих, каждая из которых характеризуется своими значениями указанных параметров. Каждая такая составляющая называется **гармоникой**.

В результате разложения речевого сигнала на входящие в его состав гармонические колебания получается частотный **спектр** – амплитудно-частотная зависимость. Например, если для сигнала частота и амплитуда первой гармоники равны F_1 и A_1 соответственно, а второй – F_2 и A_2 , то получим амплитудно-частотную зависимость, представленную на рис. 22.

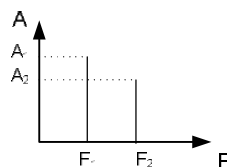


Рисунок 2 2 . Представление сигнала в виде частотного спектра

Для более сложного сигнала амплитудно-частотная зависимость будет включать гораздо большее количество вертикальных отрезков с разными значениями частоты и амплитуды. В этом случае используется другое, более наглядное представление указанной зависимости в виде огибающей кривой линии.

Спектр речевого сигнала выражает его частотный состав, т.е. самыми главными характеристиками речевого сигнала являются значения частот его гармонических составляющих. Именно в результате наложения этих частот и их восприятия органами слуха мы получаем возможность слышать богато украшенный звук – речевой сигнал. Для разложения речевого сигнала в спектр, как правило, используется преобразование Фурье [2, 3, 4, 5] (*Рылов А.С.2003кн-Анализ_Р; Рабинер Л.Р..1981кн-Цифро_О_Р_С; Фант Г.1964кн-Акуст_Т_Р; Радзишевский А.2006кн-Основ_А_и_Ц_3*). На практике такое разложение речевого сигнала представляется в виде так называемых **динамических спектрограмм** (сонограмм).

Сонограмма – это трехмерное изображение динамики артикуляторных процессов в речевом тракте. По оси абсцисс откладывается время, по оси ординат – значение частот, а степень их зачернения соответствует энергии (амплитуде) частотных компонент в спектре сигнала (рис. 23).

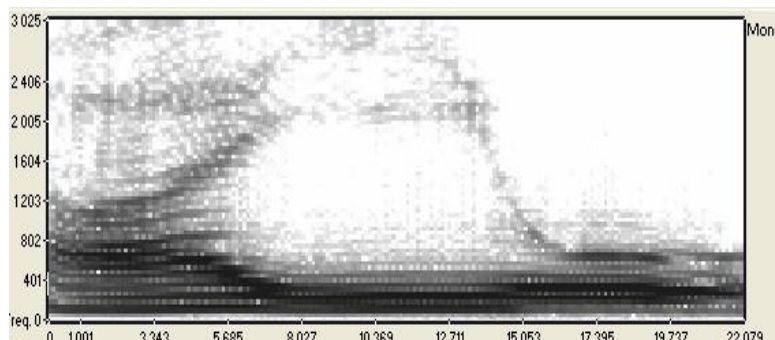


Рисунок 2 3 . Трехмерное изображение динамической спектрограммы речевого сигнала

Благодаря описанным колебательным процессам происходит образование речевого сигнала. Для изучения речевого сигнала необходимо его представить в виде набора некоторых физических и математических составляющих, которые можно изобразить в виде графиков соответствующих функций. Основными видами изображения (описания) речевого сигнала являются:

- **спектр** – разложение амплитудно-частотных зависимостей;
- **осциллограмма** – амплитудно-временная зависимость.

Представление речевого сигнала в виде осциллограммы можно увидеть в любом редакторе звука, например, с помощью любых встроенных средств записи звука на компьютере, а также в широко распространенных более мощных редакторах, таких как Sound Forge (рис. 24).



Рисунок 2 4 . Фрагмент осциллограммы фонемы [а]
(спектр фонемы представлен на рис. 23)

С точки зрения лингвистики в речевом сигнале отражаются фонетические и просодические аспекты речи, с точки зрения акустики – акустические характеристики фонем и просодем, а информационная структура содержит дополнительную информацию о речевом сигнале, включающую его смысловое содержание, а также индивидуальные характеристики личности говорящего и передающей среды.

Для реализации речевого интерфейса необходимо предварительно тщательно изучить характеристики речевого сигнала и выявить закономерности. Полученные результаты в дальнейшем используются для моделирования (синтеза) речи, а также способствуют более качественному его распознаванию. Более подробно характеристики речевого сигнала описаны в [1] (*Лобанов Б.М..2006кн-Речев_И_ИС*).

Часть 1. «Изучение фонетических закономерностей речевого сигнала»

1.1. Общие сведения о программе Sound Forge

Для изучения фонетических закономерностей речевого сигнала предлагается использовать программу *Sonic Foundry Sound Forge* (<http://www.sonicfoundry.com>) (рис. 1.1), которая имеет развитое меню, наглядную панель инструментов и удобную для работы систему окон. Программа *Sound Forge* позволяет:

- записывать с микрофона речевой сигнал и воспроизводить его осциллограмму;
- сохранять записанный речевой сигнал в файл;
- воспроизводить осциллограмму (или ее часть) ранее записанного речевого сигнала;
- изменять при необходимости масштабы просмотра осциллограммы;
- вырезать ненужные участки речевого сигнала;
- вырезать и перемещать между собой различные участки сигнала; • увеличивать и уменьшать громкость речевого сигнала; • осуществлять различные преобразования сигнала.



Рисунок 1 . 1 . Общий вид внешнего интерфейса программы Sonic Foundry Sound Forge

Рассмотрим некоторые возможности программы *Sound Forge* более подробно.

Для того чтобы записать речь и сохранить ее в файл, необходимо подключить к компьютеру микрофон, настроить его с помощью средств *Microsoft Windows* (см. разд. I) и воспользоваться кнопкой *Record* в панели инструментов *Sound Forge* (рис. 1.1). После этого откроется диалоговое окно *Record* для записи звука (рис. 1.2), в котором можно установить все необходимые атрибуты звукозаписи (в частности, частоту дискретизации, разрядность, количество каналов записи), отследить и настроить с помощью световых индикаторов уровень записи, записать речь с микрофона, а также прослушать сделанную только что запись.

Для того чтобы записать сигнал, необходимо в рассматриваемом диалоговом окне нажать на кнопку *Record* (см. рис. 1.2) и начать произносить фразу в микрофон. В процессе записи

звука почти все инструменты рассматриваемого диалога деактивируются. Для завершения записи достаточно нажать на кнопку *Stop* этого же диалогового окна.



Рисунок 1. 2. Диалоговое окно для записи звука в Sound Forge

После остановки звукозаписи ее сразу же можно прослушать, используя стандартную кнопку *Play* для воспроизведения звука, находящуюся справа от кнопки *Stop*, в этом же диалоговом окне. Если при воспроизведении окажется, что запись недостаточно качественная, то нажав на кнопку *New*, можно задать частоту дискретизации и разрядность для новой записи (рис. 1.3) и после этого записать речевой сигнал заново.

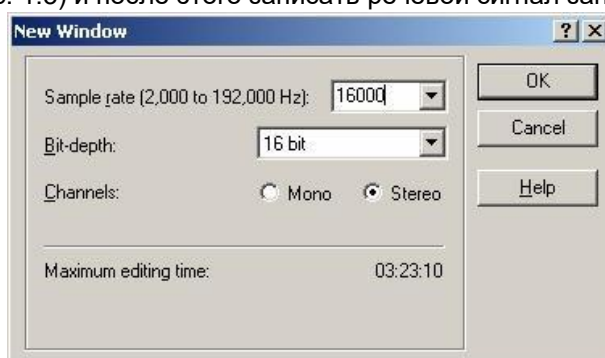


Рисунок 1.3. Установка частоты дискретизации и разрядности

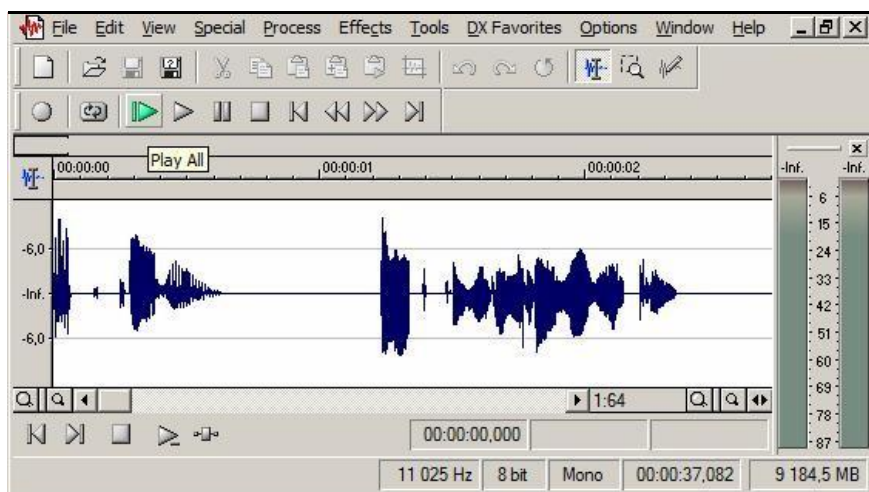


Рисунок 1 . 4 . Осциллограмма речевого сигнала в окне Sound Forge

Для того чтобы получить возможность работать с записанным речевым сигналом далее и увидеть его осциллограмму, необходимо закрыть окно *Record*. После этого откроется дочернее окно *Sound Forge*, содержащее изображение осциллограммы только что сделанной записи (рис. 1.4).

Находясь в данном окне, можно исследовать полученную запись, а также осуществлять всевозможные преобразования записанного речевого сигнала. Прежде чем приступить к исследованиям, рекомендуется сохранить запись в звуковой файл в формате wav. Для этого используются стандартные средства сохранения файлов либо на панели инструментов, либо из пункта *File* главного меню программы *Sound Forge*. Следует также отметить, что с помощью стандартного набора инструментов можно открывать файлы с ранее сохраненными звукозаписями, создавать новые, а также вырезать, копировать, перемещать фрагменты записей, выделяя их с помощью мыши прямо на осциллограмме. Для удобства дальнейшей работы выведенное на экран изображение осциллограммы можно масштабировать с помощью инструмента *Zoom*. Тогда осциллограмма принимает вид, похожий на изображение, представленное на рис. 1.5.

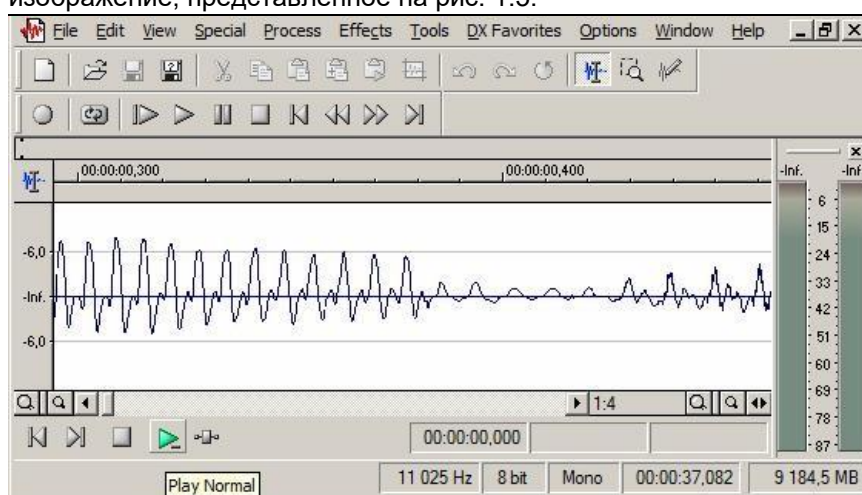


Рисунок 1 . 5 . Масштабированное изображение осциллограммы речевого сигнала

Для того чтобы прослушать записанный речевой сигнал, осциллограмма которого изображена в окне, можно использовать кнопки воспроизведения, проматывания и т.п. на панели инструментов, которые активизируются для этого окна.

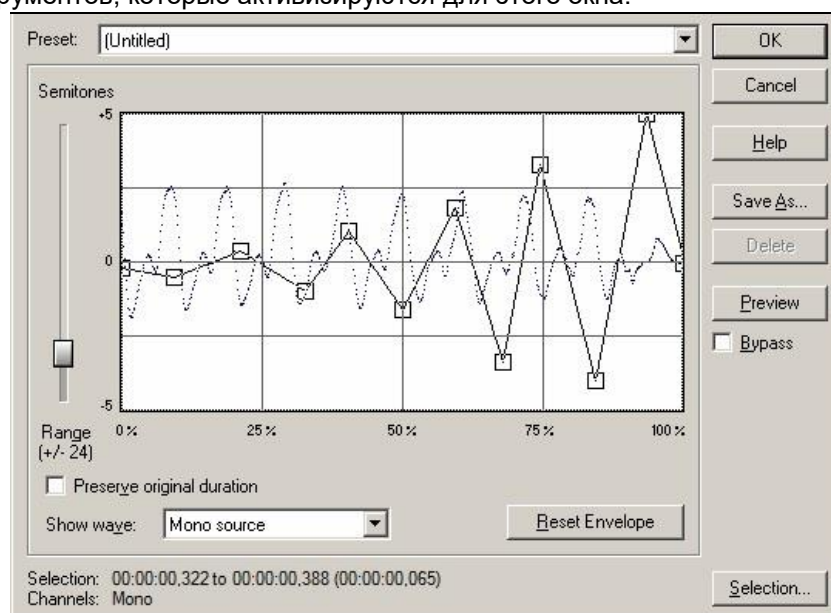


Рисунок 1 . 6 . Управление частотой основного тона речевого сигнала

В программе *Sound Forge* имеются различные способы преобразования записанного звука. В частности, для преобразования речевого сигнала иногда требуется использовать возможности изменения высоты тона аудиоданных, т.е. изменять частоту основного тона F_0 . Для этого предназначена функция *Pitch Bend*, которую можно найти с помощью меню *Effect -> Pitch -> Bend* (рис. 1.6). Например, с помощью этой функции можно сделать так, что звук будет медленно повышаться от начала до конца записи.

Обычно, когда изменяется высота тона аудиоданных, их длина также меняется. Если увеличить высоту тона, то и данные станут короче, и наоборот, – если уменьшить высоту тона, то данные станут длиннее. Именно это происходит при использовании функции *Pitch Bend*. Чтобы избежать этого, можно использовать функцию *Pitch Shift* программы *Sound Forge* для изменения высоты тона без изменения длины аудиоданных, которую можно найти с помощью меню *Effects -> Pitch -> Shift*.

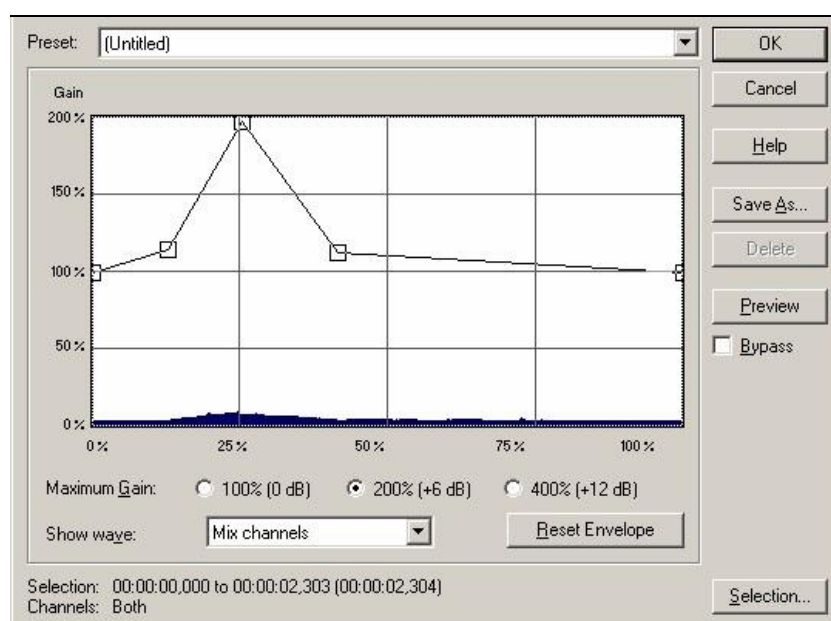


Рисунок 1 . 7 . Регулирование уровня громкости (амплитуды) речевого сигнала

Для регулирования уровня громкости (амплитуды) речевого сигнала и/или отдельных его фрагментов предназначена команда меню *Process -> Fade -> Graphic*, при обращении к которой открывается диалоговое окно *Graphic Fade* (рис. 1.7). В этом окне изображен график. На левой оси графика указываются значения амплитуды, которые могут варьироваться от 0 до 400 % (в зависимости от значения параметра *Maximum Gain*, находящегося под графиком). На графике изображена кривая, показывающая изменение громкости звуковых данных. Левый край этой кривой представляет начало выделенной области, а правый край – ее конец. Изменяя форму кривой, можно регулировать значение амплитуды различных частей выделенного фрагмента речевого сигнала.

Аналогичным образом регулируется панорамирование звука, т.е. то, как будет слышен звук в звуковом поле между двумя динамиками. Для изменения панорамирования используется команда меню *Process -> Pan/Expand*.

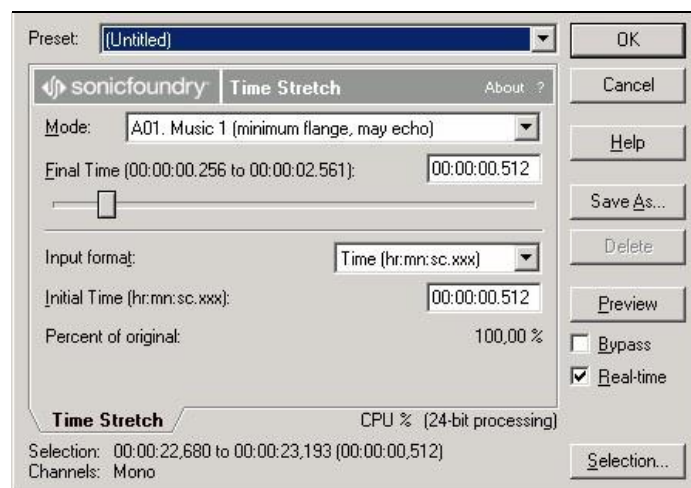


Рисунок 1 . 8 . Регулировка длины аудиоданных

В процессе обработки звука при изменении длины аудиоданных высота тона тоже изменяется. Иными словами, если укоротить данные, то высота тона увеличится, а если удлинить данные, то высота тона уменьшится. Для того чтобы избежать изменения высоты тона, в программе *Sound Forge* имеется функция, позволяющая изменить длину аудиоданных без воздействия на высоту тона. Эта функция называется *Time Compress/Expand* и вызывается из меню *Process -> Time Stretch* (рис. 1.8).

Существуют также и другие возможности преобразования звука в программе *Sound Forge*. Однако для выполнения приведенных в данном пособии лабораторных работ описанных выше средств вполне достаточно.

1.2. Фонетический состав русской речи

Фонема — это минимальная значимая звуковая единица речи [1] (*Лобанов Б.М., 2006* кн *Речев_И_ИС*). Любое сообщение можно написать буквами, а можно — фонемами.

Акустические свойства каждой фонемы обусловлены артикуляторными особенностями ее образования в речевом тракте человека, к которым относятся:

- место образования;
- способ образования.

Определение 1 . 1 . Под **местом образования** понимается сужение артикуляторного тракта, определяющее его артикуляторную статику и резонансные свойства. Место образования гласных фонем обусловлено положением тела языка (переднее / заднее, высокое / низкое) и губ (огубленное / неогубленное). Место образования согласных фонем определяет положение смычки или щели (губное, зубное, альвеолярное, нёбное, фарингальное), а также заднее или переднее положение тела языка (разделение на мягкие и твердые согласные).

Определение 1 . 2 . От **способа образования** фонем зависят энергетические и динамические особенности артикуляции фонемы. Под способом образования фонемы понимается то, каким образом образуется звук при прохождении воздуха через артикуляторный тракт человека во время произнесения фонемы.

В соответствии с артикуляторными особенностями образования, фонемы различаются:

- 1) по способу образования;
- 2) по месту образования.

По способу образования фонемы подразделяются на следующие виды (более подробно см.

[1] (*Лобанов Б.М..2006кн-Речев_И_ИС*)):

- гласные: [y], [o], [a], [э], [ы], [и];
- плавные: [й];
- боковые (латеральные): [л] (твёрдый) и [л'] (мягкий);
- носовые: [м], [м'], [н], [н'];
- дрожащие: [р], [р'];
- звонкие: [й], [л], [л'], [м], [м'], [н], [н'], [р], [р'], [з], [з'], [ж], [в], [в'], [б], [б'], [д], [д'], [г], [г'];
- глухие: [с], [с'], [ш], [ш'], [ф], [ф'], [х], [х'], [п], [п'], [т], [т'], [к], [к'];
- щелевые (фрикативные): [с], [с'], [ш], [ш'], [ф], [ф'], [х], [х'], [з], [з'], [ж], [в], [в'];
- взрывные: [п], [п'], [т], [т'], [к], [к'], [б], [б'], [д], [д'], [г], [г'];
- аффрикаты: [ц], [ч'].

Гласные отличаются по **месту образования** положением тела языка в координатах верхний / нижний – передний / задний и положением губ в координатах огубленный / неогубленный:

- верхний, задний, огубленный – [y];
- нижний, задний, огубленный – [o];
- нижний, задний, неогубленный – [a];
- нижний, передний, неогубленный – [э];
- верхний, задний, неогубленный – [ы]; • верхний, передний, неогубленный – [и].

Согласные по **месту образования** отличаются положением смычки или щели (губное, зубное, альвеолярное, заальвеолярное, нёбное), а также задним или передним положением тела языка (мягкие и твердые): • губные твёрдые – [м], [в], [б], [ф], [п];

- зубные твёрдые – [н], [л], [д], [р], [з], [с], [т], [ц];
- заальвеолярные твёрдые – [ж], [ш];
- нёбные твёрдые – [г], [к], [х];
- губные мягкие – [м'], [в'], [б'], [ф'], [п'];
- зубные мягкие – [н'], [л'], [д'], [р'], [з'], [с'], [т'];
- заальвеолярные мягкие – [ш'], [ч']; • нёбные мягкие – [й], [г'], [к'], [х'].

Полный перечень 42 фонем русского языка:

[y] [o] [a] [э] [и] [ы]

[й] [л] [л'] [м] [м'] [н] [н'] [р] [р'] [п] [п'] [к] [к'] [т] [т']

[в] [в'] [з] [з'] [ж] [б] [б'] [д] [д'] [г] [г']

[ф] [ф'] [с] [с'] [ш] [ш'] [х] [х'] [ц] [ч']

1.3. Порядок выполнения лабораторной работы - часть 1

Задание 1.1. Повторите основные характеристики речевого сигнала и основы фонетики русской речи, классификацию фонем (см. подразд. 1.2).

Задание 1.2. Научитесь работать с программой *Sound Forge* для записи речевых сигналов и отрегулируйте уровни записи и воспроизведения этих сигналов (см. подразд. 1.1). Для этого выполните следующие описанные далее действия.

1. Запустите программу *Sound Forge*. Изучите содержание меню программы, основные команды, назначение кнопок на панели инструментов.
2. Научитесь записывать речевые сигналы с помощью программы *Sound Forge*. Для этого нажмите на панели инструментов кнопку с точкой *Record*. На экране появится окно

с заголовком *Record-Sound 1*. В этом окне нажмите кнопку *Record* и отчетливо произнесите в микрофон любую фразу. После этого нажмите кнопку *Stop* и закройте

окно кнопкой *Close*. На экране появится осциллограмма произнесенной фразы. Прослушайте записанную фразу, нажав на кнопку *Play All* на панели инструментов.

3. Отрегулируйте уровень громкости воспроизведения речевого сигнала с помощью стандартной программы системы *Windows* «*Регулятор громкости*» (см. разд. I).

Для вызова этой программы 2 раза щелкните кнопкой мыши на значке *Громкость*



в правом углу панели задач *Windows*. Установите нормальный уровень громкости воспроизведения, прослушивая записанный речевой сигнал.

4. Отрегулируйте уровень записи речевого сигнала с помощью программы «*Регулятор громкости*». Для этого из режима «*Громкость*» этой программы перейдите в режим «*Запись*» (в меню «*Параметры*» выберите команду «*Свойства*» и отметьте настройку уровня записи). Уровень записи будет установлен правильно, если при записи речевого сигнала с помощью программы *Sound Forge* амплитуда сигнала на осциллограмме не будет сильно превышать значение, отмеченное зеленой горизонтальной линией.

Задание 1.3. С помощью программы *Sound Forge* исследуйте влияние частоты дискретизации и разрядности речевого сигнала на его разборчивость и качество записи. Для этого выполните следующие описанные далее действия.

1. Откройте в программе *Sound Forge* окно *Record* и нажмите кнопку *New...*
2. Установленное значение частоты дискретизации (*Sample rate*) 16 000 Гц обеспечивает среднее качество записи речевого сигнала, типичное для микрофона. Запишите произвольную фразу, прослушайте и оцените разборчивость и качество звучания. Уменьшите значение частоты дискретизации до 4 000 Гц, произнесите ту же фразу и запишите новый речевой сигнал. Прослушайте речевой сигнал и отметьте заметное ухудшение качества и разборчивости речи.
3. Установите значение частоты дискретизации 8 000 Гц, что соответствует телефонному качеству и разборчивости звучания. Запишите речевой сигнал при этом значении частоты дискретизации, прослушайте его и оцените качество звучания. Увеличьте значение частоты дискретизации до 32 000 Гц. Запишите и прослушайте речевой сигнал. Такая частота дискретизации обеспечивает высокое качество записи и звучания, но в то же время требуется значительно больший объем памяти для преобразования речевого сигнала.
4. Установленное значение разрядности речевого сигнала *Bit-Depth* 16 бит соответствует высокому качеству звучания. Установите низкую разрядность 8 бит. Запишите в том же порядке фразы при разной частоте дискретизации, прослушайте речевой сигнал и сделайте выводы об изменении качества звучания и разборчивости речи.

Задание 1.4. С помощью программы *Sound Forge* исследуйте фонетические закономерности речевого сигнала.

Задание 1.4.1. Изучите гласные фонемы русской речи. Для этого выполните следующие описанные далее действия.

Запишите гласные звуки с помощью программы *Sound Forge* следующим образом.

1. Откройте окно *Record*. В этом окне нажмите кнопку *Record* и отчетливо произнесите в микрофон гласные звуки: [а], [о], [у], [э], [ы], [и].
2. Нажмите кнопку *Stop* и закройте окно кнопкой *Close*. На экране появится осциллограмма произнесенных звуков.
3. Проверьте их звучание, нажав на кнопку *Play All* на панели инструментов. Сохраните речевой сигнал с гласными фонемами в файле *sample1.wav* с помощью команды *Save* из пункта меню *File*.

4. Найдите на осциллограмме речевого сигнала участки, соответствующие каждой гласной фонеме (участки выделяются нажатием правой кнопки мыши, озвучиваются – нажатием кнопки *Play* на панели инструментов).
5. Прослушайте, как звучит каждая гласная фонема в отдельности.
6. Исследуйте осциллограмму фонемы [а]. Для этого увеличьте масштаб осциллограммы до 1:2 с помощью кнопки *Zoom In*, находящейся в правом нижнем углу окна осциллограммы. Найдите область фонемы [а] и приблизительно зарисуйте общий вид осциллограммы в виде огибающего контура. После этого визуально определите периодичность сигнала фонемы [а] и выделите с помощью мыши один период. Прослушайте, как он звучит. Это будет похоже на щелчок. Теперь нажмите кнопку *Play Looped* (циклический повтор отмеченного сигнала), и вы услышите подобие произнесенного вами звука [а]. Для остановки нажмите кнопку *Stop*.
7. Таким же образом исследуйте и зарисуйте примерный вид осциллограмм всех остальных гласных фонем русского языка.

В результате проведенных исследований сделайте вывод о том, что гласные фонемы речи могут разборчиво звучать отдельно от других фонем, и более того, даже один повторяющийся период сигнала этих фонем звучит вполне приемлемо.

8. Запишите с помощью программы *Sound Forge* гласные «я», «ё», «ю», «е» и проведите аналогичные исследования, в результате которых покажите, какие фонемы образуют указанные буквы.

Завершите выполнение этого задания нажатием кнопки *Close Window*.

Задание 1.4.2. Изучите **шипящие** согласные фонемы русской речи. Для этого выполните следующие описанные далее действия.

1. Запишите с помощью программы *Sound Forge* следующие фонемы: [ф], [с], [ш], [х]. Сохраните этот речевой сигнал в файле *sample2.wav*.
2. На полученной осциллограмме речевого сигнала найдите участки, соответствующие каждой фонеме. Выделите с помощью мыши и отдельно прослушайте каждую фонему.
3. Выделите и отдельно прослушайте участки шипящих согласных фонем. В результате сделайте вывод о том, что шипящие согласные можно выделять и прослушивать отдельно.
4. Запишите с помощью программы *Sound Forge* следующие слоги: «фа», «са», «ша», «ха». Сохраните этот речевой сигнал в файле *sample3.wav*.
5. На полученной осциллограмме речевого сигнала найдите участки, соответствующие каждому слогу. Выделите и отдельно прослушайте слоги с шипящими согласными.
6. Выделите и отдельно прослушайте участки шипящих согласных фонем в слогах. В результате сделайте вывод о том, что шипящие согласные, в отличие от других согласных, можно выделять и прослушивать отдельно от гласных звуков. Однако обратите внимание на влияние гласных на их звучание, сравнив результаты этапов 1–3 с текущими результатами.
7. Увеличьте масштаб осциллограммы до 1:2 и найдите осциллограммы шипящих согласных фонем. Зарисуйте их примерный вид.
8. Теперь попытайтесь выделить отдельный период осциллограммы любой шипящей согласной. Нажмите кнопку *Play Looped* для циклического повторения звучания выделенного периода. Вы услышите звук, даже отдаленно не напоминающий шипящую согласную фонему.

Сделайте вывод о том, что в шипящих согласных не звучат отдельные периоды сигнала в циклическом режиме.

Завершите выполнение этого задания нажатием кнопки *Close Window*.

Задание 1.4.3. Исследование фонетических характеристик **взрывных, носовых** и всех остальных согласных русской речи.

1. Запишите с помощью программы *Sound Forge* следующие фонемы: [б], [д], [м], [н]. Сохраните этот речевой сигнал в файле *sample4.wav*.
2. На полученной осциллограмме речевого сигнала найдите участки, соответствующие каждой фонеме. Попробуйте выделить и отдельно прослушать участки каждой из согласных фонем.
3. Запишите с помощью программ *Sound Forge* следующие слоги: «ба», «да», «ма», «на» и аналогичные слоги с другими согласными. Сохраните этот речевой сигнал в файле *sample5.wav*.
4. На полученной осциллограмме речевого сигнала найдите участки, соответствующие слогам «ба» и «да». Попробуйте выделить и отдельно прослушать участки взрывных согласных фонем. Определенных звуков [б] и [д] вы не услышите, так как они как бы влились в соседний гласный звук [а].
5. В слогах «ма» и «на» попробуйте выделить и прослушать участки носовых согласных фонем. Результат будет тем же. Носовые согласные влились в соседний звук [а] и отдельно не звучат.
6. В результате сделайте вывод о том, что все согласные звуки (кроме шипящих) звучат только в контексте с гласными, т.е. происходит наложение артикуляции одного звука на другой (так называемый эффект **коартикуляции**).
7. Проанализируйте аналогичным образом все остальные согласные фонемы русской речи.
8. Увеличьте масштаб осциллограммы до 1:2 и найдите осциллограммы согласных фонем. Зарисуйте их примерный вид.
9. Теперь попытайтесь выделить отдельный период осциллограммы любой из этих согласных. Нажмите кнопку *Play Looped* для циклического повторения звучания выделенного периода.

Сделайте вывод о том, что во всех согласных фонемах (включая шипящие) не звучат отдельные периоды сигнала в циклическом режиме.

Завершите выполнение этого пункта нажатием кнопки *Close Window*.

Задание 1.5. Сравните полученные вами результаты с результатами своих однокурсников и проанализируйте влияние голосовых особенностей и особенностей произношения различных дикторов на внешний вид осциллограмм и звучание фонем.

Задание 1.6. Включите в отчет по работе выполненные вами зарисовки и сделанные в процессе работы выводы.

Примечание . Постарайтесь сохранить полученные при выполнении данной работы результаты, для того чтобы их можно было использовать при выполнении следующих работ.

Часть №2 «Изучение просодических закономерностей речевого сигнала»

2.1. Просодические характеристики речевого сигнала

Просодика описывает **ударение** и **интонацию** в речи (рис. 2.1) [1] (Лобанов Б.М., 2006 г. *Речев. И. ИС*).

Естественной речи присуще использование интонации, так как без этого речь будет звучать монотонно и неестественно. Речь без интонации очень трудно воспринимать, так как вместе с интонацией человек передает дополнительный смысл своих высказываний. Аналогично естественной чертой речи является также наличие ударений. Нам понятно использование ударений в отдельных словах, так как без ударений невозможно будет отличить одно слово от другого и понять смысл произносимых сообщений. Кроме словесных ударений в просодике речи рассматриваются также другие виды ударений, которые делают естественную речь более выразительной и в том числе помогают передавать ее интонацию.



Рисунок 2 . 1 . Просодические характеристики речи

Из рис. 2.2 видно, что акцентная группа (АГ) делится на три составляющие: **предъядро**, **ядро** и **заядро**. Это так называемые элементы акцентной группы (ЭАГ). При этом ядро считается главным элементом АГ. Правила разбивки акцентной группы на элементы достаточно просты: **ядро** является главной ударной гласной слова, **заядро** – это все звуки справа от ядра, **предъядро** – все звуки слева от ядра. Получается, что заядро и предъядро могут содержать различное количество звуков вплоть до их отсутствия. Например, фонетический период: «Ты поедешь сегодня? – Да!» – содержит фразу «Да!», которая является при этом и синтагмой, и акцентной группой одновременно. В этом случае единственный гласный звук придется делить на три части, потому что нужно выделить интонационный тип и определить ядро, заядро и предъядро.

F0

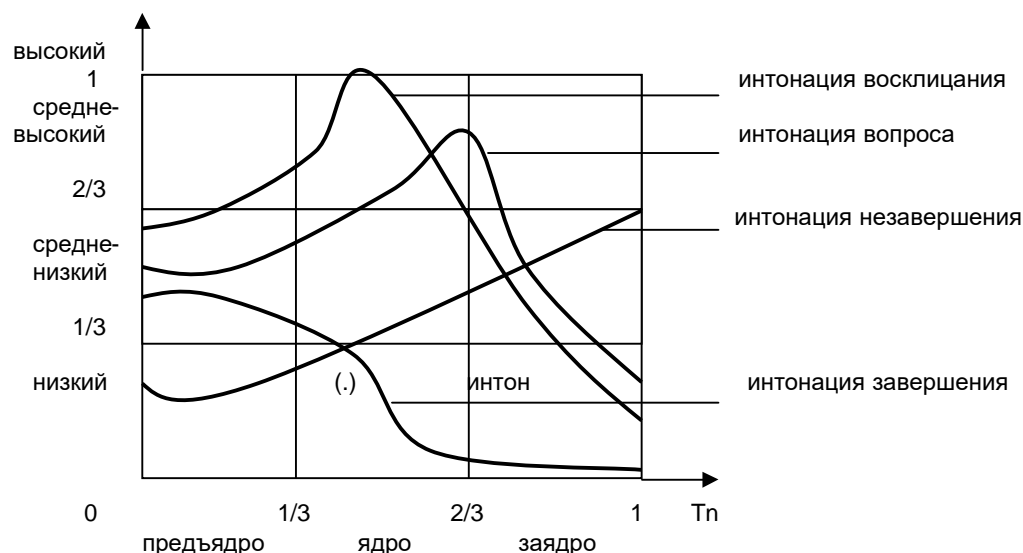


Рисунок 2.2. Мелодические контуры четырех интонационных типов конечной акцентной группы

Интонационный тип фразы определяется положением мелодического максимума относительно главного ударного гласного. Причем положению максимума в центре гласного соответствует интонация побуждения или восклицания, сдвигу его влево – интонация завершения или утверждения, а сдвигу вправо – интонация незавершенности или вопроса. Другие параметры мелодической кривой (значение максимума и ее ширина) существенны при передаче эмоциональных оттенков речи.

2.2. Порядок выполнения лабораторной работы - часть №2

Задание 2.1. Повторите основные понятия просодики. Вспомните основные фонетические особенности речевого сигнала.

Задание 2.2. Исследуйте просодические закономерности речевого сигнала.

Задание 2.2.1. Исследуйте явление ударения в просодике и понятие акцентной группы.

1. Запишите с помощью программы *Sound Forge* слово «Саша». Сохраните его в файле *sample2_1.wav*.
2. Запишите с помощью программы *Sound Forge* слово «кушал». Сохраните его в файле *sample2_2.wav*.
3. Запишите с помощью программы *Sound Forge* слово «кашу». Сохраните его в файле *sample2_3.wav*.
4. Создайте новый файл и перенесите в него записанные ранее слова в следующих последовательностях (для этого каждое слово нужно вначале выделить и вырезать с помощью кнопки *Cut* на панели инструментов, а затем вставить в нужном месте осциллограммы нажатием кнопки *Paste*):
 - «Саша», «кушал», «кашу»; •
 - «кушал», «Саша», «кашу»; •
 - «кушал», «кашу», «Саша»; •
 - «Саша», «кашу», «кушал»; •
 - «кашу», «Саша», «кушал»; •
 - «кашу», «кушал», «Саша».
5. Прослушайте полученные фразы и попытайтесь выявить интонационные различия в них. Сделайте вывод о том, что все они звучат с точки зрения интонации примерно одинаково и включают по 3 акцентные группы, что можно записать следующим образом: (+) (+) (+).

6. Сохраните результат в файле *sample2_4.wav*.
7. Запишите с помощью программы *Sound Forge* фразу «Саша кушал кашу» целиком с ударением на первое слово. Сохраните ее в файле *sample2_5.wav*.
8. Прослушайте фразу – она звучит как утверждение. Теперь рассматриваемая фраза состоит из одной акцентной группы вида (+ – –).
9. Найдите на осциллограмме речевого сигнала участок, соответствующий слову «Саша». Переставьте это слово в конец фразы.
10. Прослушайте измененную фразу – теперь она звучит почти как вопрос. При этом акцентная группа преобразовалась к виду (– – +).
11. Сохраните ее в файле *sample2_6.wav*.
12. Проведите различные эксперименты, меняя местами слова фразы. Сравните полученные результаты с результатами пунктов 1–5.
13. Еще раз произнесите фразу, но с ударением не на первое слово, а на второе или третье. Снова поменяйте местами слова фразы, прослушайте, что у вас получилось, и сравните с предыдущими результатами.
14. Придумайте самостоятельно (или прочитайте из любого текста) несколько различных фраз и попытайтесь выделить в них акцентные группы. Обратите внимание на то, что более длинные и сложные фразы могут состоять из нескольких акцентных групп.

Задание 2.2.2. Исследуйте явление интонации в просодике.

1. Запишите с помощью программы *Sound Forge* фразу «Саша кушал кашу» с различными видами интонации:
 - простая завершенность (которая обозначается на письме с помощью точки (.));
 - незавершенность ((,) и (–)) – для этого попробуйте продолжить фразу, например, словами «сегодня на обед»;
 - вопрос ((?)); • восклицание ((!)).
 Сохраните записанные фразы либо в один файл, либо в разные файлы с расширением **.wav*.
2. Прослушайте записанные фразы и сравните их. Нарисуйте соответствующие каждой фразе интонационные контуры и расстановку ударений.
3. Проведите аналогичный эксперимент с целью анализа эстетической и эмоциональной функций просодики (например, прочитайте фразу с выражением и без, с чувством гордости, радости, огорчения и т.п.).
4. Придумайте самостоятельно (или прочитайте из любого текста) несколько различных фраз и проведите аналогичные исследования.

В результате проведенных исследований сделайте вывод о влиянии просодических характеристик (интонации и ударения) на речевой сигнал.

Задание 2.3. Научитесь работать с мелодическими контурами.

1. Запишите с помощью программы *Sound Forge* фразу «Студент писал реферат» монотонно, т.е. на одном тоне и без выделения какого-либо слова. Скопируйте записанную фразу либо после записанной, либо в новый аудиофайл и сохраните обе записи.
2. Выберите в программе *Sound Forge* режим *Effect -> Pitch -> Bend*. Подбирая мелодический контур изменения F_0 , добейтесь звучания интонации завершенности для первой записи, соответствующей фразовому ударению:
 - на первое слово (ответ на вопрос «Кто писал реферат?»);
 - на второе слово (ответ на вопрос «Что делал студент?»);
 - на третье слово (ответ на вопрос «Что писал студент?»).
3. Запишите полученные звуковые файлы. Зарисуйте общий вид полученных интонационных контуров F_0 .
4. Подбирая контур изменения F_0 , добейтесь звучания интонации вопроса во фразе «Студент писал реферат» с ударениями на первое, второе и третье слово.

5. Запишите полученные звуковые файлы. Зарисуйте общий вид полученных интонационных контуров F_0 .
6. Используя в программе *Sound Forge* режим *Process -> Fade -> Graphic*, увеличьте силу звука (амплитуду) ударных слов в записанных файлах. Добейтесь более отчётливо выраженного фразового ударения.
7. Запишите полученные звуковые файлы. Зарисуйте общий вид полученных контуров амплитуды A .
8. Используя в программе *Sound Forge* режим *Process -> Time Stretch*, увеличьте длительность ударных слов в записанных файлах. Добейтесь более отчётливо выраженного фразового ударения.
9. Проведите аналогичные преобразования для второй записи фразы «*Студент писал реферат*». При этом вместо режима *Effect -> Pitch -> Bend* используйте режим *Effects -> Pitch -> Shift*.
10. Запишите полученные звуковые файлы.
11. Сравните проведенные преобразования для обеих записей одной и той же фразы. Сделайте соответствующие выводы.

Задание 2.4. Включите в отчет по работе выполненные вами зарисовки и сделанные в процессе работы выводы.

Часть №3 «Исследование компиляционных методов синтеза речевого сигнала»

3.1. Методы синтеза речи

Все существующие в настоящее время методы синтеза речевого сигнала делятся на два основных класса [1] (*Лобанов Б.М..2006кн-Речев_И_ИС*):

- параметрические;

Параметрические методы синтеза речевого сигнала характеризуются тем, что при их реализации используются знания о том, каким образом осуществляется образование речи у человека и какими параметрами описывается речевой сигнал. Основываясь на этих знаниях, осуществляется моделирование акустических процессов речеобразования. В результате такого моделирования речевой сигнал генерируется подобно тому, как это делается в речевом тракте человека. Моделирование процессов речеобразования человека в системах автоматического синтеза речи осуществляется одним из двух способов:

- на основе знаний об артикуляторных процессах, происходящих в речевой системе человека;
- путем математического описания собственно акустических процессов.

Компиляционные методы синтеза речевого сигнала игнорируют физику процесса речеобразования. В них для синтеза речи используются отрезки естественной речевой волны, и из них в дальнейшем собирается (компилируется) речевой сигнал. В качестве отрезков речевой волны могут использоваться самые разные сегменты:

- микроволны (отрезок, вырезаемый из записей отдельных фонем или аллофонов);
- аллофоны;
- дифоны (всевозможные паросочетания звуков) и др.

Каждый из указанных классов методов синтеза речи обладает своими достоинствами и недостатками.

Компиляционные методы имеют следующие достоинства:

- синтезированная речь звучит близко к естественной;
- с помощью компиляционного метода можно «научить» компьютер разговаривать голосом любого человека (например своим голосом) – это качество используется в активно развивающемся в настоящее время направлении клонирования голоса [6, 7] (*Лобанов Б.М.2002ст-Комп_К_П_Г; Цирульник Л.И.2006ст-Автом_С_К*).

Недостатки компиляционных методов:

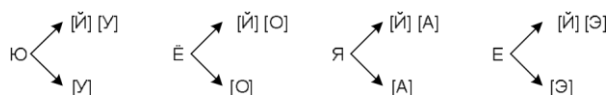
- трудоемкость разработки;
- необходимость довольно больших объемов памяти для хранения баз голосов.

3.2. Правила транскрибирования

Как указано выше, основной идеей компиляционных методов синтеза речи является использование сегментов естественной речевой волны для компиляции из них слитных потоков речи. Однако в процессе преобразования естественно-языкового текста в речь предварительно необходимо преобразовать этот текст в последовательность фонем, из которых фактически состоит речевой поток.

Для преобразования орфографического текста в фонемный используются правила транскрибирования. В каждом языке приняты свои правила произношения слов, которым соответствуют свои правила транскрибирования. Рассмотрим здесь некоторые правила транскрибирования, характерные для русской речи.

В русском языке 10 гласных букв: У, О, А, Э, Ы, И, Ю, Ё, Я, Е. Десяти гласным буквам ставятся в соответствие 6 гласных фонем: [У], [О], [А], [Э], [Ы], [И]. Получается, что 4 гласные буквы Ю, Ё, Я и Е как бы «лишние». Они превращаются либо в пары фонем, либо в отдельные фонемы следующим образом:



При этом используются следующие правила транскрибирования [8] (Печенева Т.А.2004сост-Виды_Я_Р):

1. «Превращение гласной буквы Ю, Ё, Я или Е в пару фонем [Й] и [У], [О], [А] или [Э] соответственно происходит в тех случаях, когда эти буквы в слове стоят в одном из следующих положений: в начале транскрибируемого слова; после мягкого знака; после твердого знака; после гласной буквы». Например: выюга -> [в й у г а], объём -> [а б й о м], яма -> [й а м а], поёт-> [п а й о т].
2. «В остальных случаях после согласных гласные Ю, Ё, Я или Е преобразуются в фонемы [У], [О], [А] или [Э] соответственно и при этом смягчают стоящие перед ними согласные». Например: тюк -> [т' у к], лёд -> [л' о т], пять -> [п' а т'], семь -> [с' э м'].

Из этих правил имеются исключения, в частности, в тех случаях, когда рассматриваемая гласная является безударной. В этом случае звук изменяется (редуцируется). Например, безударная буква Я может быть преобразована в звук [И]: тянуть -> [т' и н у т'].

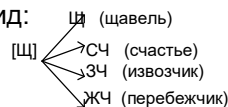
В русском языке есть правило, согласно которому безударная гласная буква О преобразуется в фонему [А], а безударная Е – в [И]. Например: молоко -> [м а л а к о], дерево -> [д' э р и в а].

Еще правило: «Буква И после согласных Ж и Ш преобразуется в звук [Ы]». Например: жить -> [ж ы т'], шить -> [ш ы т'].

Как видно из формулировок, указанные правила хорошо описываются алгоритмически.

Имеются различные правила преобразования согласных. Рассмотрим некоторые из них:

1. Правило образования звука [Щ] имеет следующий вид:



Ц (царь)

2. Фонема [Ц] образуется одним из следующих двух способов:

ТЬС (купаться)



3. Правило оглушения: звонкая согласная перед глухой согласной или в конце слова преобразуется в парную ей глухую. Например: лёд -> [л' о т], буг -> [б у к].
4. Правило озвончения: если глухая согласная стоит перед звонкой, то звук преобразуется в звонкий. Например: косьба -> [к а з' б а], просьба -> [п р о з' б а].
5. Для некоторых пар согласных работает явление уподобления. Например: сшить -> [шит'].
6. Смягчение парных согласных: если после согласной стоит мягкая согласная, то и эта согласная также смягчается. Например: хвостик -> [х в о с' т' и к].

Здесь также есть исключения: слово «бог» преобразуется в транскрипцию [б о х], а не в [б о к], слово «что» преобразуется в [ш т о].

Все эти правила хорошо известны из фонетики [9] (Зиндер Л.Р.1979кн-Общая_Ф) и легко поддаются алгоритмизации.

3.3. Порядок выполнения лабораторной работы – часть №3

Задание 3.1. Повторите классификацию методов синтеза речевого сигнала. Повторите понятие системы автоматического синтеза речи. Вспомните или изучите обобщенную структуру системы синтеза речи [1] (Лобанов Б.М..2006кн-Речев_И_ИС). Изучите правила транскрибирования, используемые в русской речи.

Задание 3.2. Запишите несколько фраз естественного языка на бумаге, а затем прочитайте их в микрофон и запишите в файлы *.wav в программе *Sound Forge*. Эти фразы будем считать эталонными. Они вам понадобятся для дальнейшего сравнительного анализа. При этом желательно предусмотреть наличие фраз следующего типа:

- фраза из одной синтагмы;
 - фраза из нескольких синтагм;
 - фразы с различным типом интонации;
 - фразы с перечислениями, т.е. сложные предложения с запятыми, союзами и т.п.;
- общее количество фраз не должно быть слишком большим – примерно от 3 до 5.

Задание 3.3. Практическое исследование компиляционных методов синтеза речевого сигнала.

Задание 3.3.1. Подготовьте фонетическую базу для синтеза речи.

1. Для выполнения данного задания возьмите записанные вами при выполнении лабораторной работы №1 файлы с гласными и согласными фонемами (если они у вас не сохранились, то вам придется записать их заново).
2. Запишите на бумаге выбранные вами при выполнении задания 2 фразы естественного языка в виде фонемного текста, т.е. в виде последовательности фонем.

Задание 3.3.2. Сформируйте звуковые файлы на основе фонемных текстов.

1. Создайте новый пустой файл в программе *Sound Forge*.
2. На основе подготовленных ранее фонемных текстов сформируйте соответствующие фразы речи, компилируя их из подготовленных наборов фонем русского языка. Например, если у вас записано фонемное слово [с а ш а], то в пустой файл скопируйте сначала осциллограмму фонемы [с], затем – [а], [ш] и [а].
3. Таким же образом запишите в отдельные файлы все выбранные вами фразы.

Задание 3.3.3. Проанализируйте скомпилированную речь.

1. Прослушайте «синтезированные» вами речевые высказывания и сравните их с эталонными. Сравните их осциллограммы.
2. Сделайте выводы о степени разборчивости и естественности звучания. Отметьте в отчете основные положительные и отрицательные стороны компиляционного метода синтеза речи.

Задание 3.3.4. Попытайтесь улучшить качество звучания записанной речи путем дополнительной обработки осциллограммы.

1. Выберите какую-нибудь одну фразу.
2. Посмотрите еще раз внимательно на осциллограмму эталонной фразы и синтезированной. Попробуйте путем выбрасывания лишних участков осциллограммы, склеивания, перетаскивания, увеличения или уменьшения амплитуды, длительности и других подобных действий добиться более качественного звучания.
3. Сделайте соответствующие выводы.

Задание 3.3.5. Попытайтесь улучшить качество звучания записанной речи путем совершенствования фонетической базы.

1. Разбейте подготовленные вами фонемные тексты на отдельные слоги и запишите эти слоги аналогично тому, как были надиктованы отдельные фонемы.
2. Осуществите компиляцию речи для своих фраз с использованием заготовленных слогов.
3. Сравните полученные фразы с эталонными и с теми, которые были сформированы ранее. Сделайте соответствующие выводы. Отметьте, какие фонетические и просодические особенности речи следует учитывать при разработке синтезаторов речи.

Задание 3.3.6*. Попытайтесь улучшить естественность звучания записанной речи путем обработки контуров мелодики, ритмики и энергетики.

1. Вспомните просодические параметры речи.
2. Зарисуйте мелодические, ритмические и энергетические контуры синтезируемых фраз.

3. Попробуйте описать сформированные контуры путем перечисления приблизительных или относительных значений просодических параметров $F_0(t)$, $A(t)$, $T(t)$.
4. На основе описанных параметров попробуйте сделать синтезированные вами фразы более интонационно насыщенными, используя любые известные вам методы обработки просодических параметров. Сделайте соответствующие выводы.

Задание 3.4. Включите в отчет по работе выполненные вами зарисовки и сделанные в процессе работы выводы, представьте все рабочие материалы (естественно-языковые фразы, фонемные тексты, просодически размеченные тексты и т.д.). Изобразите обобщенную структуру синтезатора речи, который использует компиляционный метод синтеза, промоделированный при выполнении данной работы. Данный вид синтезаторов иногда называют компиляторами.

Сделайте общий вывод по работе, ответив на вопрос, какие особенности речи следует учитывать для того, чтобы получить более качественное и естественное звучание. Каких компонентов недостает для того, чтобы разработанный вами «синтезатор» стал полноценной системой автоматического синтеза речи с неограниченным словарем?

Литература

1. *Лобанов Б.М. 2006кн-Речев_И_ИС*
Лобанов, Б. М. Речевой интерфейс интеллектуальных систем : учеб. пособие / Б. М. Лобанов, О. Е. Елисеева ; под науч. ред. В. В. Голенкова . – Минск : БГУИР, 2006. – 152 с.
2. *Рылов А.С. 2003кн-Анализ_Р*
Рылов, А. С. Анализ речи в распознающих системах / А. С. Рылов. – Минск : Бестпринт, 2003. – 264 с.
3. *Рабинер Л.Р. 1981кн-Цифро_О_Р_С*
Рабинер, Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер ; пер. с англ. – М. : Радио и связь, 1981. – 495 с.
4. *Фант Г. 1964кн-Акуст_Т_Р*
Фант, Г. Акустическая теория речеобразования / Г. Фант ; пер. с англ.; под ред. В. С. Григорьева. – М. : Наука, 1964. – 284 с.
5. *Радзишевский А. 2006кн-Основ_А_и_Ц_3*
Радзишевский, А. Основы аналогового и цифрового звука / А. Радзишевский. – М. : Вильямс, 2006. – 288 с.
6. *Лобанов Б.М. 2002ст-Комп_К_П_Г*
Лобанов, Б. М. Компьютерное «клонирование» персонального голоса и речи / Б. М. Лобанов // Новости искусственного интеллекта. – 2002. – №5. – С. 35–39.
7. *Цирульник Л.И. 2006ст-Автом_С_К*
Цирульник, Л. И. Автоматизированная система клонирования фонетико-акустических характеристик речи / Л.И. Цирульник // Информатика. – 2006. – № 2(10). – С. 46–55.
8. *Печенева Т.А. 2004сост-Виды_Я_Р*
Виды языкового разбора : краткий комплексный словарь-справочник / Авт.-сост. Т. А. Печенева. – Минск : Изд. центр БГУ, 2004. – 99 с.
9. *Зиндер Л.Р. 1979кн-Общая_Ф*
Зиндер, Л. Р. Общая фонетика / Л. Р. Зиндер. – М. : Высш. шк., 1979. – 312 с.