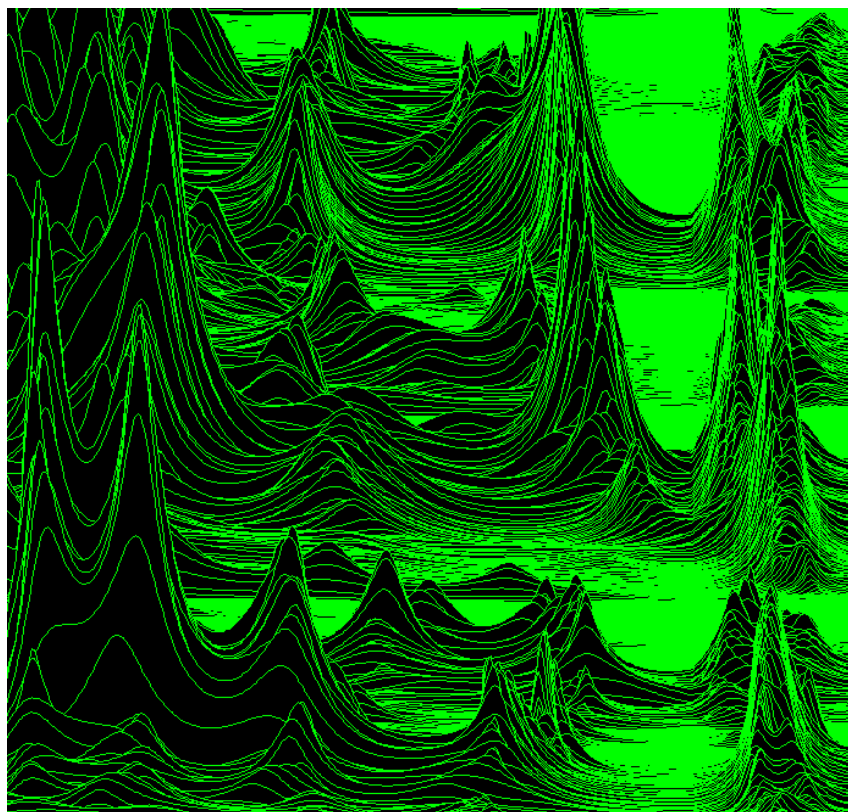


М.Б. Столбов, А.-Р. Кассу

ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ



Санкт-Петербург
2016

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
УНИВЕРСИТЕТ ИТМО

М.Б. Столбов, А-Р. Кассу
ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ
Учебно-методическое пособие
по лабораторному практикуму



Санкт-Петербург

2016

Столбов М.Б., Кассу А.-Р. М. Цифровая обработка речевых сигналов: Учебно-методическое пособие по лабораторному практикуму – СПб: НИУ ИТМО, 2016. – 71 с.

Пособие содержит краткое описание лабораторного практикума по методам цифровой обработки речевых сигналов, используемым в системах анализа и синтеза речевых сигналов, а также системах автоматического распознавания речи и системах автоматической идентификации и верификации дикторов.

Пособие адресовано студентам, обучающимся по специальности 09.04.02 и направлениям подготовки «Речевые информационные системы» и «Системы голосового самообслуживания».

Рекомендовано к печати Ученым советом факультета информационных технологий и программирования от 27.10.2015 №10

Примечание: задания по лабораторным работам рассчитаны на использование примеров – звуковых файлов, которые должны сопровождать данное учебно-методическое пособие. Звуковые файлы находятся на компакт-диске.

Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в



2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2016

© М. Б. Столбов, А.-Р. М. Кассу, 2016

Содержание

Часть 1. МЕТОДЫ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ.....	4
ЛАБОРАТОРНАЯ РАБОТА № 1 Основы работы с программой SIS II.....	4
ЛАБОРАТОРНАЯ РАБОТА № 2 Спектральные характеристики	7
ЛАБОРАТОРНАЯ РАБОТА № 3 Автокорреляционная функция.....	12
ЛАБОРАТОРНАЯ РАБОТА № 4 Спектрограмма	15
ЛАБОРАТОРНАЯ РАБОТА № 5 Кепстр.....	21
ЛАБОРАТОРНАЯ РАБОТА № 6 Гистограмма и частота пересечения нуля ..	27
ЛАБОРАТОРНАЯ РАБОТА № 7 Отношение сигнал-шум и время реверберации.....	30
СПИСОК ЛИТЕРАТУРЫ.....	33
 Часть 2. АЛГОРИТМЫ КОМПЕНСАЦИИ ИСКАЖЕНИЙ РЕЧЕВЫХ СИГНАЛОВ.....	 34
ЛАБОРАТОРНАЯ РАБОТА № 1 Основы работы с программой Sound Cleaner II.....	34
ЛАБОРАТОРНАЯ РАБОТА № 2 Импульсные помехи и динамическая обработка сигналов.....	39
ЛАБОРАТОРНАЯ РАБОТА № 3 Частотная фильтрация	42
ЛАБОРАТОРНАЯ РАБОТА № 4 Фильтрация широкополосного шума	47
ЛАБОРАТОРНАЯ РАБОТА № 5 Выделение речи на фоне тональных помех	50
ЛАБОРАТОРНАЯ РАБОТА № 6 Деклиппирование и дереверберация.....	53
ЛАБОРАТОРНАЯ РАБОТА № 7 Выравнивание спектра	57
ЛАБОРАТОРНАЯ РАБОТА № 8 Подавление помех, вызванных работой аппаратов связи.....	60
ЛАБОРАТОРНАЯ РАБОТА № 9 Выделение речи с использованием алгоритмов стереофильтрации.....	64
ЛАБОРАТОРНАЯ РАБОТА № 10 Комплексное применение фильтров	68
СПИСОК ЛИТЕРАТУРЫ	68
ПРИЛОЖЕНИЕ 1 Пример оформления титульного листа	70
ПРИЛОЖЕНИЕ 2 Рекомендации по выполнению лабораторных работ.....	71

Часть 1. МЕТОДЫ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ

ЛАБОРАТОРНАЯ РАБОТА № 1 Основы работы с программой SIS II

Цель работы: ознакомление с основами работы с программой SIS II и оценкой характеристик сигналов.

Основные сведения по работе с программой содержатся в руководстве пользователя [1], а также в разделе HELP меню программы SIS II.

Краткие теоретические сведения

Цифровой сигнал – аналоговый сигнал, дискретизованный по времени и квантованный по амплитуде.

Аналого-цифровое преобразование непрерывного сигнала включает дискретизация во времени и квантование по амплитуде.

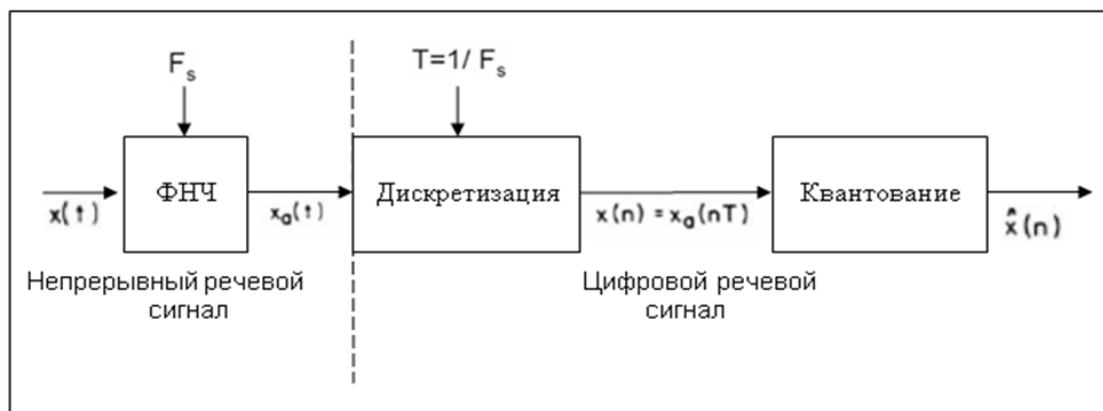


Рисунок 1 – Схема аналого-цифрового преобразования сигнала

На рисунке обозначены: F_s – частота дискретизации сигнала, T – шаг дискретизации, $\text{ФНЧ} = F_s/2$. Например, $F_s = 16 \text{ кГц} \leftrightarrow \text{ФНЧ} = 8 \text{ кГц}$.

Квантование сигнала по амплитуде — 8-, 16-, 24-, 32-разрядное (например, при 8-разрядном квантовании амплитуды цифрового сигнала находятся в интервале $[+127 \dots -128]$).

Дискретизация непрерывного аналогового сигнала через равные промежутки (шаг дискретизации) T : $x(n) = X(nT)$.

Частота дискретизации измеряется в герцах (отсчетах в секунду). Частота дискретизации F_s связана с шагом дискретизации по времени T следующим соотношением: $F_s = 1/T$.

Например:

$$F_s = 8000 \text{ Гц} \leftrightarrow T = 1/8000 = 125 \text{ мкс}$$

$$F_s = 10000 \text{ Гц} \leftrightarrow T = 1/10000 = 100 \text{ мкс}$$

$$Fs = 20000 \text{ Гц} \leftrightarrow T = 1/20000 = 50 \text{ мкс}$$

$$Fs = 40000 \text{ Гц} \leftrightarrow T = 1/40000 = 25 \text{ мкс}$$

Выборочные оценки параметров сигналов: дисперсии, мощности, энергии и СКО (в окне в N отсчетов)

Оценка среднего значения:

$$\mu x = \frac{1}{N} \sum_{i=1}^N x(i) = \langle x(i) \rangle$$

где $\langle \rangle$ - оператор усреднения по времени.

$$\text{Центрирование процесса: } x_o(i) = x(i) - \langle x(i) \rangle = x(i) - \mu x$$

Оценка дисперсии:

$$Dx = \langle x_o^2(i) \rangle = \frac{1}{N} \sum_{i=1}^N x^2(i) - \left[\frac{1}{N} \sum_{i=1}^N x(i) \right]^2$$

$$\text{СКО (RMS): } \sigma x = [Dx]^{1/2}$$

Свойство дисперсии: если сигнал является суммой сигналов независимых источников, тогда его дисперсия равна сумме дисперсий этих сигналов:

$$Dx + Dy = \langle (x(i) + y(i))^2 \rangle = \langle x(i)^2 + 2x(i)y(i) + y(i)^2 \rangle = \langle x(i)^2 \rangle + \langle y(i)^2 \rangle = Dx + Dy$$

Оценка мощности (среднеквадратического значения):

$$Px = \langle x^2(i) \rangle / Tw = \frac{1}{N} \sum_{i=1}^N x^2(i) = Dx + \mu x^2 = \sigma x^2 + \mu x^2$$

Оценка энергии в окне N:

$$Ex = \sum_{i=1}^N x^2(i)$$

Наиболее распространенной формой представления цифровых сигналов в звуковых редакторах является осциллограмма. Осциллограмма – это представление цифрового сигнала в координатах «амплитуда-время». Эффекты дискретизации и квантования находят свое проявление в осциллограмме.

Задание 1

Загрузите файл **ST_rms1000.wav** (стереосигнал, «белый шум»). Разделите стереосигнал на два сигнала моно. Поделите один из сигналов на 1000, а за-

тем умножьте на 1000. Сравните осциллограммы сигналов и объясните результат.

Задание 2

Загрузите файл *rms1000.wav*. Вычислите «энергию» (с окном 20 мс) и постройте график в отдельном окне.

Увеличьте (или уменьшите) амплитуду сигнала вдвое. Как зависит оценка энергии от амплитуды сигнала?

Добавьте к исходному сигналу константы 2000, 4000, вычислите энергию и постройте графики полученных сигналов в том же окне.

Объясните полученные результаты. По какой формуле вычисляется «энергия» в программе SIS II?

Задание 3

Загрузите файл *16kHz_rms2000.wav* («белый шум», 16 кГц, СКО 2000).

Вычислите энергию и постройте график в отдельном окне.

Передискретизируйте сигнал на 8 кГц, вычислите энергию и постройте график полученного сигнала в том же окне.

Передискретизируйте сигнал обратно с 8 кГц на 16 кГц, вычислите энергию и постройте график полученного сигнала в том же окне.

Как изменилась энергия сигнала?

Объясните полученные результаты.

Задание 4

Загрузите файлы *8kHz_sin.wav*, *16kHz_sin.wav* (синусоидальный сигнал, частота 3,9 кГц, амплитуда 1000, частоты дискретизации 8 кГц и 16 кГц).

Постройте осциллограммы в одном окне.

Чем различаются осциллограммы сигнала при разных частотах дискретизации?

Задание 5

Загрузите файл *rms1000.wav* («белый шум», СКО 1000).

Вычислите энергию с размерами окна 4 мс, 20 мс, 100 мс.

Постройте графики в одном окне.

Как зависит оценка энергии от размера окна?

ЛАБОРАТОРНАЯ РАБОТА № 2

Спектральные характеристики

Цель работы: исследование свойств оценок спектра Фурье для широкополосных и узкополосных сигналов, применение спектров для анализа характеристик тональных сигналов, представление спектров в линейном масштабе и логарифмическом масштабе.

Краткие теоретические сведения

Спектральное представление сигналов, оценка спектра

Дискретный спектр – это представление дискретной функции $x(k)$ на интервале $[0, N - 1]$ в виде суммы периодических компонент [2, 3]:

$$\begin{aligned}x(k) &= DFT\{Ax(i)\} \\x(k) &= \sum_{i=0}^{N-1} Ax(i) \exp[-j2\pi ik/N + \Phi(i)] = \\&= \sum_{i=0}^{N-1} Ax(i) [\cos(2\pi ik/N + \Phi(i)) - j \sin(2\pi ik/N + \Phi(i))] \\&= \sum_{i=0}^{N-1} [Re(i) + jIm(i)] [\cos(2\pi ik/N) - j \sin(2\pi ik/N)] \\&= \sum_{i=0}^{N-1} X(i) \exp[-j2\pi ik/N]\end{aligned}$$

где i – индекс частоты, соответствующий частоте $fi = i \times Fs/N$, $0 \leq i \leq N - 1$, N – размер кадра, $A(i)$ – амплитудный спектр, $X(i)$ – комплексный спектр.

Дискретный спектр сигнала вычисляется на основе следующих соотношений:

$$\begin{aligned}X(i) &= DFT\{x(k)\} = \sum_{k=0}^{N-1} x(k) \exp[-j2\pi ik/N] = \\&= \sum_{k=0}^{N-1} x(k) [\cos(2\pi ik/N) - j \sin(2\pi ik/N)]\end{aligned}$$

Комплексный спектр

$$X(i) = Re(i) + j Im(i) = |X(i)| \exp(j\pi\Phi_x(i))$$

Амплитудный спектр

$$A(i) = |X(i)|$$

Фазовый спектр

$$\Phi x(i) = \arctan[Im(i)/Re(i)]$$

Периодограмма (мгновенный спектр мощности)

$$|X(i)|^2 = Re(i)^2 + Im(i)^2$$

Периодограмма имеет смысл распределения мощности сигнала (энергии сигнала на кадре) по частотным полосам. Например, в белом шуме средний спектр равномерный, мощность сигнала распределена по частотным полосам равномерно.

Обычно на графиках показывают средний амплитудный спектр, вычисленный с помощью усреднения амплитудных спектров по времени

$$\langle Ax(i) \rangle = \langle |X(i, k)| \rangle = \frac{1}{K} \sum_{k=1}^K |X(i, k)|$$

где i – индекс частоты, k – индекс времени (кадров), K – число кадров усреднения.

Спектр в точке (мгновенный спектр) означает амплитудный спектр, вычисленный на одном кадре $Ax(i) = |X(i)|$.

Поскольку реальная часть спектра является симметричной (по частоте), а мнимая – антисимметричной, то спектр мощности и амплитудный спектр являются симметричными функциями. Поэтому на графиках показывается только половина этих спектров плюс 1 бин: $i = 0, \dots, N/2$.

Средний геометрический спектр

Средний геометрический спектр определяется на основе следующих эквивалентных соотношений:

$$G_{xx}(i) = \left[\prod_{k=1}^K |X(i, k)| \right]^{1/K} = \exp \left[\frac{1}{K} \sum_{k=1}^K \ln |X(i, k)| \right]$$

где i – индекс частоты, k – индекс времени (кадров), K – число кадров усреднения.

Средний геометрический спектр приближается на зашумленном сигнале к среднему спектру шума и не превышает средний арифметический спектр.

Децибелы

Амплитуда спектра речевых сигналов сильно меняется с частотой. Поэтому для графического представления спектров речевых сигналов часто применяют логарифмическую шкалу (децибел, дБ, dB).

Децибел – это число, выражающее в логарифмической мере ($lg = \log_{10}$) отношение R двух величин (измеряемой и эталонной); обычно употребляется при большом отношении этих величин.

Децибелы для мощностей $R = 10 \lg(P/P_0)$

Децибелы для амплитуд

$$R = 20 \lg(A/A_0) = 10 \lg(A/A_0)^2 = 10 \lg(P/P_0)$$

Децибелы для мощностей = Децибелы для амплитуд!

Примеры:

$$R = 10 \Rightarrow R_{\text{дБ}} = 20 \lg(10) = 20 \text{ дБ}$$

$$R = 100 \Rightarrow R_{\text{дБ}} = 20 \lg(100) = 40 \text{ дБ}$$

$$R = 1000 \Rightarrow R_{\text{дБ}} = 20 \lg(1000) = 60 \text{ дБ}$$

$$R_{\text{дБ}} = 60 \text{ дБ} \approx 20 \lg(2^{10}) = 10 \times 20 \lg(2) \Rightarrow 20 \lg(2) = 6 \text{ дБ} \Rightarrow R = 2 \sim 6 \text{ дБ}$$

$$3 \text{ дБ} = 1/2 \times 6 \text{ дБ} = 1/2 \times 20 \lg(2) = 20 \lg(2^{1/2}) \Rightarrow R = 1.41 \sim 3 \text{ дБ}$$

Задание 1

Исследование спектра «белого шума».

Сгенерируйте «белый шум» (или загрузите файл **White_noise_11025.wav**). Постройте в одном окне средние спектры сигнала с разными размерами кадра (1024, 512, 256).

Как влияет размер окна на оценку среднего спектра?

Объясните изменение амплитуды спектра.

Постройте в одном окне средние спектры сигнала на интервалах сигнала разной длины (спектры сигнала на интервале между временными марками).

Как влияет длина выборки на оценку среднего спектра?

Увеличьте (уменьшите) амплитуду сигнала в 2 раза. Постройте в одном окне средние спектры сигналов с разной амплитудой.

Как изменится спектр (в линейном и логарифмическом масштабе)?

Объясните изменение амплитуды спектра.

Задание 2

Исследование спектра тонального сигнала.

Сгенерируйте синус 2 кГц с частотой дискретизации 11025 Гц, длительностью 15 с, амплитудой 20000.

Постройте в одном окне средние спектры сигнала с разными размерами кадра (1024, 512, 256).

Как влияет размер окна на оценку среднего спектра?

Объясните изменение амплитуды спектра.

Как изменится спектр (в линейном и логарифмическом масштабе)?

Задание 3

Исследование оценок спектра тонального сигнала.

Сгенерируйте тональный сигнал с частотой дискретизации 11025 Гц, длительностью 15 с, амплитудой 20000, периодом 10 отсчетов (гармоника 1102,2 Гц).

Постройте в одном окне мгновенные спектры на разных участках сигнала.

Объясните различие спектров.

Постройте в одном окне средние спектры сигнала с разными размерами кадра (256, 512, 2048).

Объясните изменение оценки среднего спектра от размера кадра?

Постройте в одном окне средние спектры сигнала с разными взвешивающими окнами. Характеристики различных взвешивающих окон описаны в руководстве пользователя.

Чем различаются средние спектры одного и того же сигнала для разных типов взвешивающих окон?

Задание 4

Обнаружение гармоник с близкими частотами.

Загрузите файл *40sin1_sin2.wav*.

Постройте в одном окне средние спектры сигнала с разными размерами кадра и разными взвешивающими окнами. Добейтесь разделения спектральных пиков сигналов, используя различные размеры кадра и типы взвешивающего окна в соответствии с рекомендациями, изложенными в руководстве пользователя.

Задание 5

Обнаружение слабой гармоникой вблизи сильной.

Загрузите файл *sin1_smallsin2.wav*.

Постройте в одном окне средние спектры сигнала с разными размерами кадра и разными взвешивающими окнами. Добейтесь разделения спектральных пиков сильной и слабой гармоник, используя различные размеры кадра и типы взвешивающего окна в соответствии с рекомендациями, изложенными в руководстве пользователя.

Задание 6

Обнаружение гармоник в шуме.

Сгенерируйте тональный сигнал (с амплитудами 2, 12, 200) в белом шуме (СКО 2000, максимальное значение амплитуды $\approx 4 \times \text{СКО}$).

Можно ли обнаружить тональный сигнал на осциллограмме?

Постройте в одном окне средние спектры сигнала с разными размерами кадра.

Исследуйте возможность обнаружения слабой гармоник в зависимости от размера кадра.

Задание 7

Спектр периодического импульсного сигнала.

Загрузите файл **11_100Hz.wav**.

Определите по осциллограмме приблизительный период импульсного сигнала и рассчитайте соответствующую частоту импульсов.

Постройте в одном окне средние спектры сигнала с разными размерами кадра (256-512-1024-2048).

Постройте средний спектр, в котором отображается соответствующая частота импульсов.

Каким образом в среднем спектре проявляется структура периодического импульсного сигнала?

Задание 8

Исследование влияния частоты дискретизации на спектр тонального сигнала.

Сгенерируйте синус с частотой 15 кГц с частотами дискретизации 20050 Гц и 40100 Гц. Постройте в одном окне средние спектры сигналов. Объясните положение максимумов в спектрах.

ЛАБОРАТОРНАЯ РАБОТА № 3

Автокорреляционная функция

Цель работы: исследование свойств оценок автокорреляционной функции (АКФ).

Краткие теоретические сведения

Автокорреляция

Для центрированного сигнала $x(i)$, $i = 0, 1, \dots, N - 1$ функция автокорреляции оценивается следующим образом:

$$R_{xx}(m) = \frac{1}{N} \sum_{i=0}^{N-m-1} x(i)x(m+i), m = 0, 1, \dots, M$$

Несмещенная оценка АКФ

$$R_{xx}(m) = \frac{1}{N-m} \sum_{i=0}^{N-m-1} x(i)x(m+i)$$

Свойства АКФ:

$$R_{xx}(-m) = R_{xx}(m), R_{xx}(0) = D_x, R_{xx}(m) \rightarrow 0 \text{ при } m \rightarrow \infty$$

Если сигнал является суммой сигналов разных независимых источников, тогда его АКФ будет равна сумме АКФ этих сигналов: $R_{ss}(m) = \sum R_{xxk}(m)$

Нормализованная АКФ (energy normalized correlation)

$$r_{xx}(m) = R_{xx}(m)/R_{xx}(0)$$

Оценки АКФ зависят от следующих параметров: интервал суммирования N и интервал задержек M (максимальное запаздывание).

Средняя оценка АКФ вычисляется применением операции «усреднение» к коррелограмме.

Коррелограмма – это форма графического представления результатов кратковременного корреляционного анализа сигналов (на последовательности кадров) в координатах время-частота-яркость / время-задержка-яркость (положительные значения обозначаются темным, отрицательные – светлым, нулевые значения – серым).

Пример

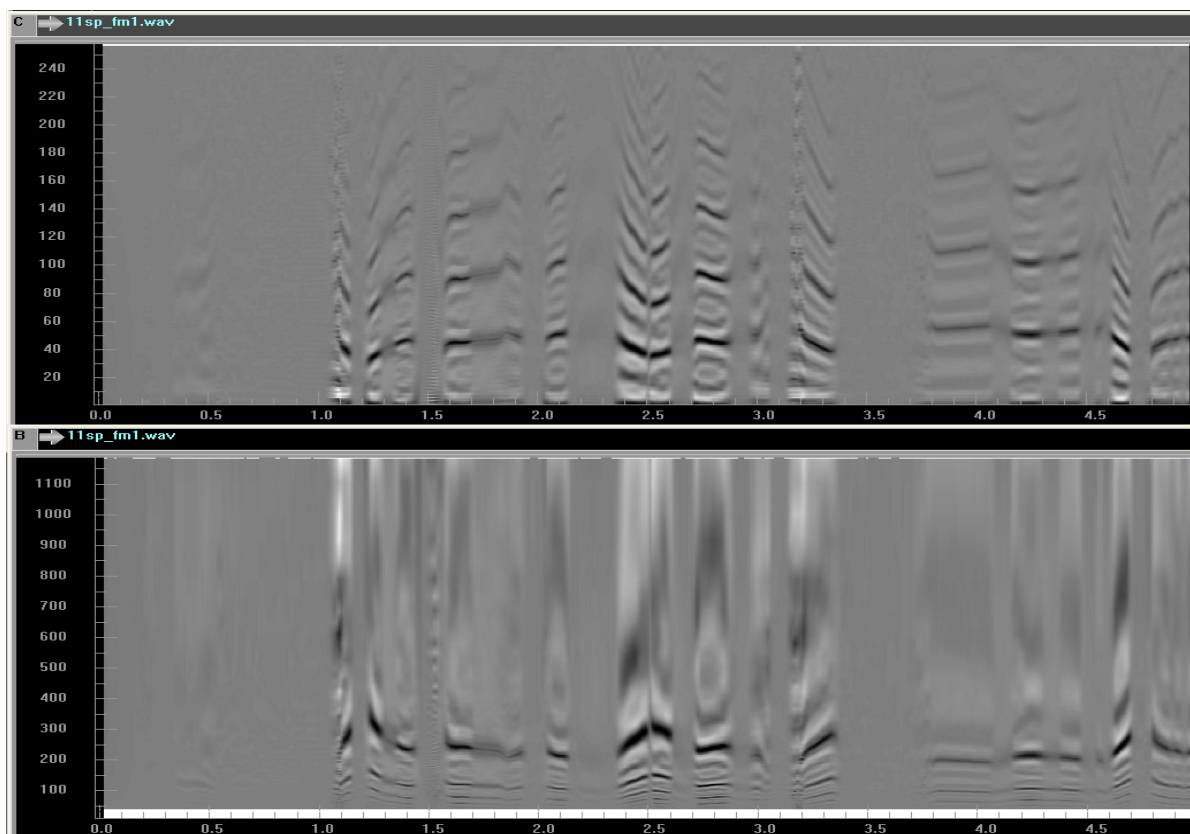


Рисунок 2 – Коррелограмма речевого сигнала:
 наверху – временное представление (время-задержка-яркость),
 внизу – частотное представление (время-частота-яркость)

Задание 1

Загрузите файл *white_noise_11025.wav* («белый шум»).

Постройте коррелограмму и средние оценки АКФ с разным размером окна (256-512-2048) и разными типами взвешивающих окон (прямоугольное и др.).

Исследуйте (и опишите) зависимость средней АКФ от размера и типа окна

Задание 2

Загрузите файл *white_noise_11025.wav*.

Постройте коррелограмму и среднюю оценку АКФ.

Увеличьте амплитуду сигнала в 2 раза. Как средняя АКФ зависит от уровня сигнала?

Добавьте к сигналу константу. Как средняя оценка АКФ зависит от величины константы? Объясните результаты.

Задание 3

Сгенерируйте периодический гармонический (тональный) сигнал, амплитудой 500, частотой дискретизации 11025 Гц, частотой тона 580 Гц (или загрузите файл *11_sin.wav*).

Постройте коррелограмму и среднюю АКФ с разным размером окна.

Как средняя АКФ зависит от размера и типа окна?

Объясните уменьшение размаха АКФ с увеличением задержки.

Можно ли наблюдать периодическую компоненту сигнала на спектрограмме, коррелограмме? На средней АКФ?

Задание 4

Сгенерируйте периодический гармонический (тональный) сигнал с частотой дискретизации 11025 Гц, амплитудой 500, частотой тона 580 Гц (или загрузите файл *11_sin.wav*). Постройте среднюю АКФ.

Сгенерируйте периодический гармонический (тональный) сигнал с частотой дискретизации 22050 Гц, амплитудой 500, частотой тона 580 Гц (или загрузите файл *22_sin.wav*). Постройте в прежнем окне среднюю АКФ.

Как изменилась средняя АКФ при увеличении частоты дискретизации сигнала вдвое?

Задание 5

Загрузите файлы *11_sin.wav* (чистый тональный сигнал) и *11_wn_sin.wav* (зашумленный тональный сигнал).

Постройте для этих сигналов средние АКФ и сравните их. Объясните различие оценок АКФ.

Задание 6

Загрузите файл *11_100Hz.wav* (периодический импульсный сигнал). Определите по осциллограмме приблизительную величину периода повторения импульсов.

Постройте среднюю АКФ и определите по ней период повторения импульсов.

ЛАБОРАТОРНАЯ РАБОТА № 4

Спектрограмма

Цель работы: исследование широкополосной и узкополосной спектрограммы Фурье (3D БПФ) и спектрограммы модели линейного предсказания (3D КЛП).

Краткие теоретические сведения

Спектрограмма – это изображение кратковременного спектра сигнала (спектров на последовательности кадров) в координатах время-частота-яркость (цвет). Интенсивность темного отображает амплитуду спектра. Чем больше интенсивность темного, тем больше Спектрограмма – самый распространенный и удобный инструмент для анализа речевых сигналов (РС) и визуализации тональной и формантной структуры РС. амплитуда. Спектрограммы строятся на основе синусоидальной модели сигнала (3D БПФ) или модели авторегрессии (3D КЛП).

В зависимости от размера окна (кадра данных) различают узкополосную и широкополосную 3D БПФ спектрограммы речевых сигналов.

Узкополосная спектрограмма

Спектральный анализ на окнах (кадрах) 50 мс.

Ширина спектральных полос анализа 40 Гц.

Шаг анализа 1 мс.

Позволяет наблюдать тональную структуру речи.

Широкополосная спектрограмма

Спектральный анализ на окнах (кадрах) 15 мс.

Ширина спектральных полос анализа 125 Гц.

Шаг анализа 1 мс.

Позволяет наблюдать импульсную и формантную структуру речи.

На рисунке 3 представлен пример узкополосной и широкополосной спектрограммы для одного и того же фрагмента речевого сигнала.

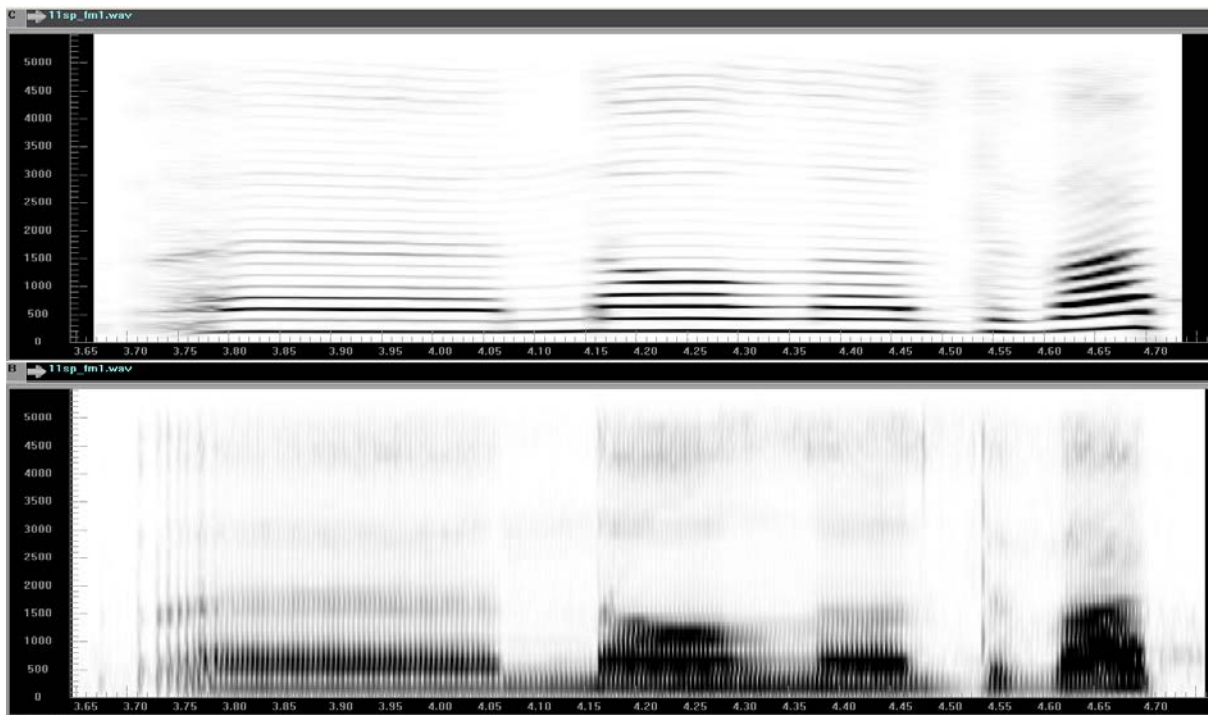


Рисунок 3 – Наверху – узкополосная спектрограмма (окно 512 отсчетов), внизу – широкополосная спектрограмма (окно 32 отсчетов) фрагмента женской речи

Другой способ визуализации (3D КЛП) формантной структуры речи основан на применении модели линейного предсказания процессов.

Модель линейного предсказания [3, 4]

Модель авторегрессии выражает текущее значение процесса через конечную линейную совокупность предыдущих значений процесса и возбуждающих импульсов $\varepsilon(t)$. Процесс авторегрессии можно интерпретировать как сигнал на выходе БИХ фильтра, на вход которого поступает широкополосный процесс (случайный белый шум или последовательность импульсов возбуждения) $\varepsilon(i)$:

$$y(i) = a_1 y(i-1) + a_2 y(i-2) + \dots + a_p y(i-p) + \varepsilon(i)$$

Коэффициенты a_i называются коэффициентами линейного предсказания (КЛП).

Оценка коэффициентов линейного предсказания

Умножим $y(i)$ на $y(i-j)$. Переходя к математическим ожиданиям величин, получим систему уравнений для функции автокорреляции:

$$R(0) = a_1 R(-1) + a_2 R(-2) + \dots + a_p R(-p) + D\varepsilon \text{ при } j = 0$$

.....

$$R(j) = a_1 R(j-1) + a_2 R(j-2) + \dots + a_p R(j-p) \text{ при } j > 0$$

Делением уравнений $R(0)$ получаем соотношения для нормализованной автокорреляционной функции:

$$\rho(j) = a_1\rho(j-1) + a_2\rho(j-2) + \dots + a_p\rho(j-p)$$

Записывая уравнение для АКФ для $j = 1, \dots, p$ и учитывая $\rho(i) = \rho(-i)$, получим систему линейных уравнений для коэффициентов a_i

$$\rho(1) = a_1 + a_2\rho(1) + \dots + a_p\rho(p-1)$$

$$\rho(2) = \dots$$

$$\dots\dots\dots$$

$$\rho(p) = a_1\rho(p-1) + a_2\rho(p-2) + \dots + a_p$$

Эта система называется системой уравнений Юла-Уокера. Решая ее, получаем оценку КЛП.

Энергетический спектр процесса авторегрессии (КЛП-спектр)

Запишем уравнение процесса авторегрессии в операторном виде [3]:

$$A(z)y(t) = \varepsilon(t)$$

Оператор $A(z)$ можно рассматривать как фильтр, после свертки которого с входными данными $y(i)$ получается временной ряд белого шума $\varepsilon(i)$. Тогда спектр сигнала $y(i)$ равен энергетическому спектру выходного сигнала (белого шума), деленному на частотную характеристику фильтра:

$$Y(z) = E(z)/A(z), z = \exp(-j2\pi f\Delta t)$$

Тогда энергетический спектр процесса авторегрессии $y(i)$ может быть представлен в следующем виде [4]:

$$S_{yy}(f) = D\varepsilon / |1 + \sum_{n=1}^p a_n \exp(j2\pi fn\Delta t)|^2$$

Последовательность спектров (спектрограмму) модели авторегрессии на последовательности кадров называют частотным откликом КЛП или 3D КЛП.

Модель авторегрессии хорошо согласуется с моделью речеобразования [4]. Максимумы энергетического спектра процесса авторегрессии соответствуют формантам (резонансам системы речеобразования). Поэтому частотный отклик КЛП является удобным инструментом визуализации формантных

траекторий. Для моделирования спектра с M формантными максимумами бывает достаточно $P = (2M + 1)$ коэффициентов. Однако, частота дискретизации и размер окон (кадров) данных также влияют на оценки КЛП-спектров. Обычно для оценки используют модель с $P = Fs/1000$ коэффициентами [4], где Fs – частота дискретизации сигнала (Гц).

Размер окон (кадров данных) обычно задается следующим образом:

$Fs < 10$ кГц: $N = 256$,

$10 \text{ кГц} \leq Fs < 20 \text{ кГц}$: $N = 512$,

$20 \text{ кГц} \leq Fs < 40 \text{ кГц}$: $N = 1024$,

$Fs > 40 \text{ кГц}$: $N = 2048$.

Ни рисунке 4 представлен мгновенный КЛП-спектр речевого сигнала на окне (кадре) данных и КЛП-спектрограмма (3D КЛП).

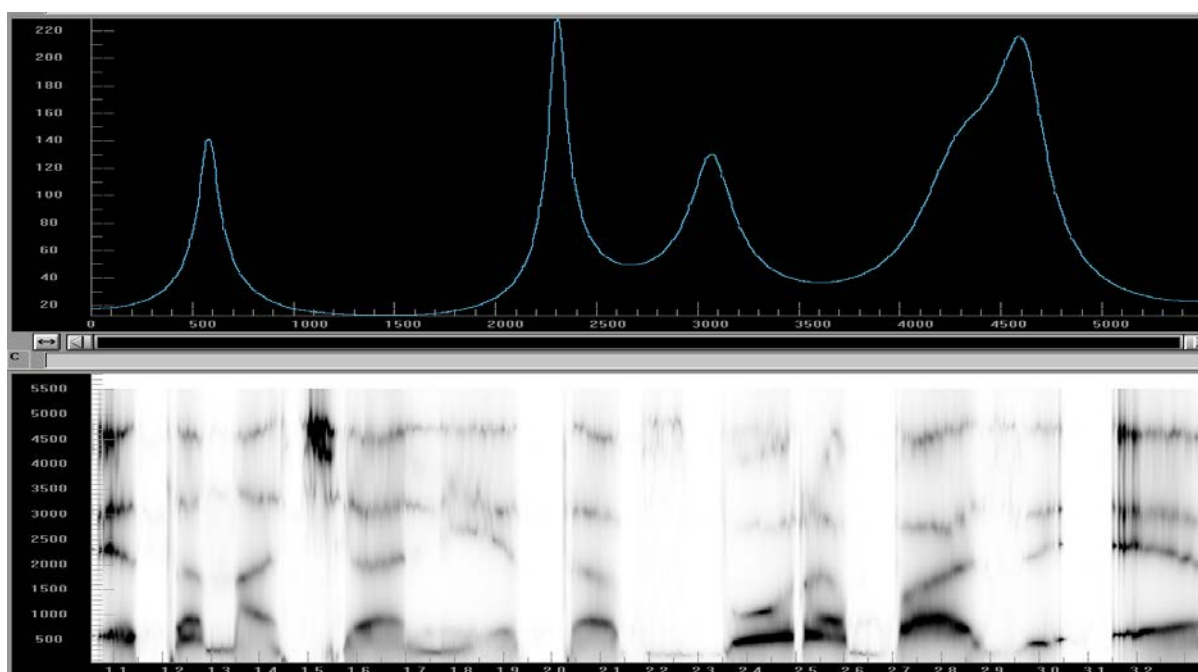


Рисунок 4 – Внизу: частотный отклик КЛП (размер кадра 128, шаг 16, число коэффициентов 12), наверху: мгновенный спектр на одном из временных срезов

Формантная траектория – последовательность значений частот максимумов спектра (формант): $F1(t), F2(t), F3(t) \dots$ Формантные траектории строятся на основе БПФ-анализа (широкополосная спектрограмма) либо авторегрессионного анализа (частотного отклика КЛП). На рисунке 5 приведены формантные траектории, построенные в программе SIS II на основе каждого из этих видов анализа.

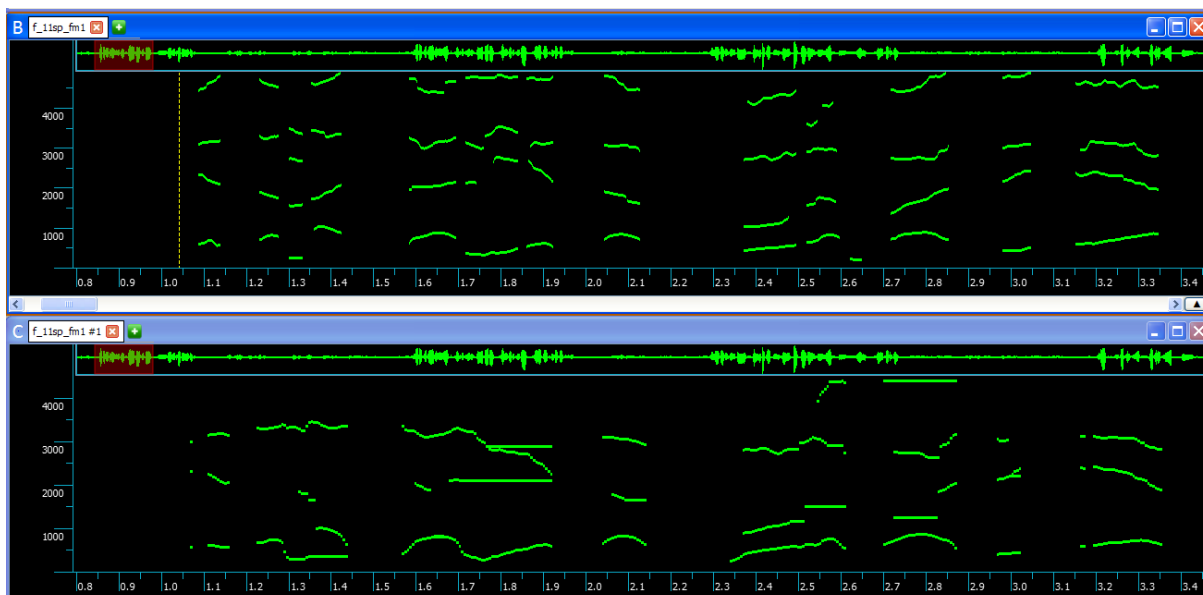


Рисунок 5 – Формантные траектории (программа SIS II) фрагмента речевого сигнала: наверху – построенные по 3D БПФ, внизу – построенные по частотному отклику 3D КЛП

Задание 1

Загрузите файл *11_100Hz.wav* (периодический импульсный сигнал).

Постройте 3D БПФ спектрограммы сигнала с различными длинами окон (32, 512, 2048).

При каких длинах окон проявляются импульсные и волновые свойства сигнала?

Попробуйте установить связь между импульсной и волновой спектрограммами.

Задание 2

Загрузите файл *11sp_fm1.wav*.

Постройте в одном окне мгновенные спектры с различными размерами кадра (32, 512, 2048).

При каком размере кадра в спектре обнаруживаются форманты речевого сигнала?

Задание 3

Загрузите файл *11sp_fm1.wav*.

Постройте 3D БПФ спектрограммы с различными длинами окон (32, 512, 2048).

При каких параметрах анализа обнаруживаются формантные траектории речевого сигнала?

Постройте 3D КЛП спектрограмму с аналогичными формантными траекториями.

Задание 4

Загрузите файл *11sp_fm1.wav*.

Постройте 3D КЛП спектрограммы с различным числом коэффициентов модели (7, 15, 20) и различными длинами окон (64, 256, 1024).

Найдите приблизительное положение формантных траекторий на 3D КЛП спектрограмме.

К чему ведет использование недостаточного/избыточного числа коэффициентов модели КЛП?

Задание 5

Загрузите файлы *11sp_m.wav*, *22sp_m.wav*, *44sp_m.wav*.

Постройте 3D КЛП спектрограммы.

Подберите количество КЛП необходимое для визуализации формантных траекторий в каждом из этих сигналов.

Задание 6

Загрузите файл *11sp_fm1.wav*.

Постройте для фрагмента (0-3,4 с) сигнала формантные траектории FFT и LPC (раздел «Модули»).

Задание 7

Исследование характеристик оценки среднего спектра от длины выборки.

Загрузите файл *white_noise_11025.wav*.

Постройте для короткого (0,25 с) и длинного (2 с) фрагмента сигнала БПФ спектрограммы с прямоугольным окном 512 и окном Хана 512.

Усредните спектрограммы и постройте оценки среднего спектра и дисперсию оценок среднего спектра (раздел «Анализ», вкладка «Усреднение»).

Постройте графики среднего спектра и дисперсии его оценки.

Объясните, как и почему меняется дисперсия оценки среднего спектра в зависимости от размера фрагмента сигнала?

ЛАБОРАТОРНАЯ РАБОТА № 5

Кепстр

Цель работы: исследование свойств кепстра и кепстрограммы и их применения для оценки основного тона (ОТ) речевых и периодических импульсных сигналов.

Краткие теоретические сведения

Определение кепстра:

$$Cx(\tau) = IFT\{Log Sxx(f)\}$$

Определение кепстра аналогично определению автокорреляционной функции:

$$Rx(\tau) = IFT\{Sxx(f)\}$$

где $Sxx(f)$ – спектр мощности сигнала $x(t)$: $Sxx(f) = |FT\{x(t)\}|^2$

Дискретный вещественный кепстр

$$Cx(m) = IDFT\{Ln|X(n)|^2\} = IDFT\{Ln Sxx(n)\}$$

где $X(n) = DFT\{X(k)\}$ – спектр сигнала на кадре k , $X(k)$ – массив временных отсчетов сигнала на кадре k .

Коэффициенты кепстра

Номера коэффициентов кепстра соотносятся с частотами (так же, как у корреляционной функции): $Fo(n) = Fs/n = 1/Tn$,

где Fs – частота дискретизации процесса, T – шаг дискретизации.

Коэффициенты с младшими номерами ($n = 1, 2, \dots$) соответствуют большим значениям частот $Fo(n)$, а с большими номерами – малым значениям частот (большим временным периодам). Таким образом, для анализа периодических импульсных сигналов с большим периодом To (малой частотой Fo) необходимо вычислять кепстры с большим окном анализа $N \times T \geq To$.

Коэффициент с нулевым индексом соответствует среднему значению логарифма спектра мощности подобно тому, как у автокорреляционной функции коэффициент с нулевым индексом соответствует среднему значению спектра мощности, то есть мощности процесса.

В целом коэффициенты кепстра имеют следующий физический смысл.

$Cx[0]$ – среднее значение логарифма спектра,

$Cx[1]$ – общий наклон спектра,

$Cx[Fs/1500] \dots Cx[Fs/500]$ – формантная структура спектра речевых сигналов,

$Cx[Fs/Fo_max] \dots Cx[Fs/Fo_min]$ – события с частотой основного тона речевых сигналов $Fo = 100$ Гц ... 500 Гц,

$Cx[Fs/80] \dots Cx[N]$ – события в пространстве частот с малым шагом неравномерности спектра (большими временными периодами периодического импульсного сигнала),

$Cx[N] \sim Fs/N$ – события с минимальной частотой периодического импульсного сигнала Fo).

Длина окна анализа определяет минимальную наблюдаемую частоту периодического сигнала (максимальный наблюдаемый период), представленную в кепстре.

$$Fmin = Fs/N \leftrightarrow Nmin = Fs/Fo$$

Например, для $Fo = 80$ Гц, $Fs = 8$ кГц имеем $Nmin = 200$!

Кепстрограмма – форма графического представления функции кепстра (от кадра к кадру в координатах «время-частота-яркость» / «время-задержка-яркость» (положительные значения обозначаются темным, отрицательные и нулевые значения – светлым).

Средний (по времени) кепстр представляется в редакторе в координатах «амплитуда-задержка». По координате времени на графике среднего кепстра даются временные задержки, соответствующие номерам коэффициентов кепстра.

Примеры кепстров речевых сигналов

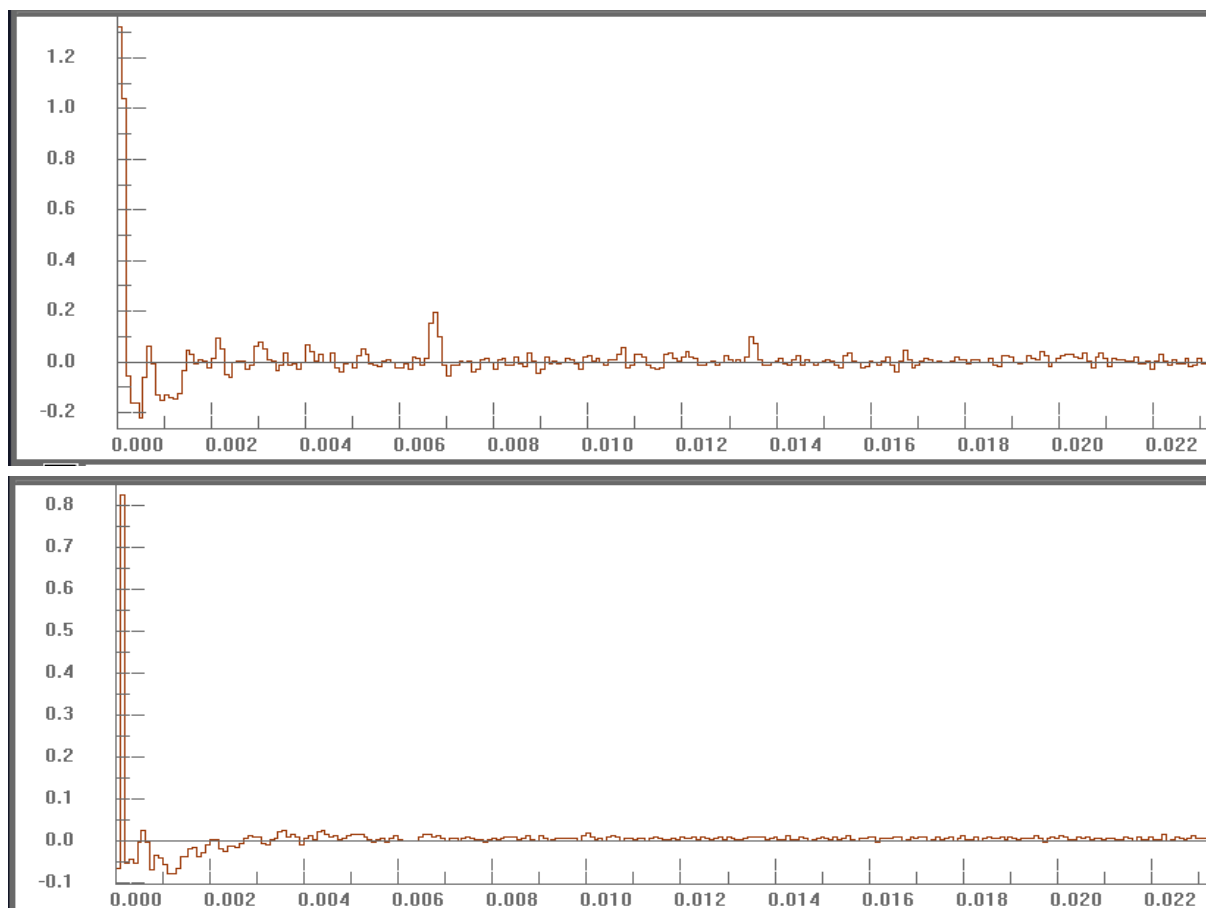


Рисунок 6 – Наверху – мгновенный кепстр тонального речевого сигнала, внизу – средний кепстр речевого сигнала на интервале 30 с

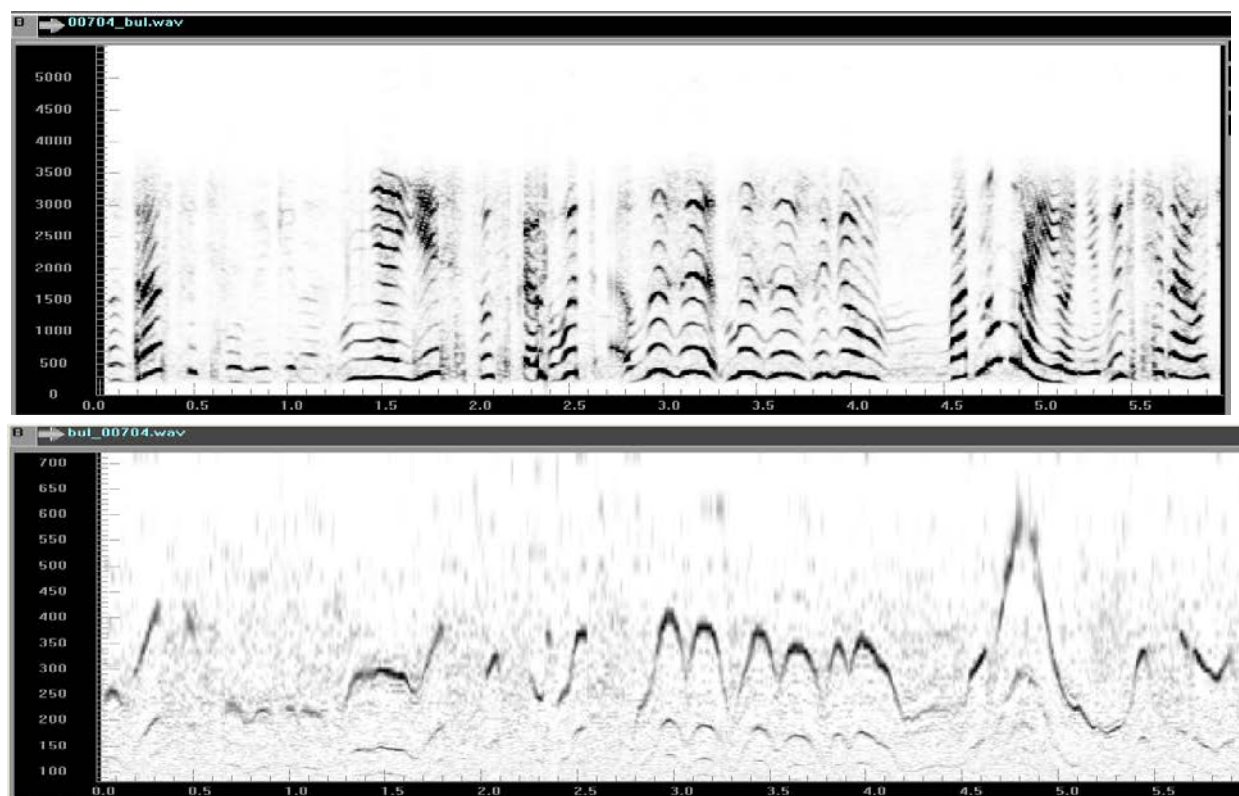


Рисунок 7 – Наверху - спектрограмма фрагмента речевого сигнала, внизу - кепстрограмма фрагмента речевого сигнала (высокие частоты основного тона)

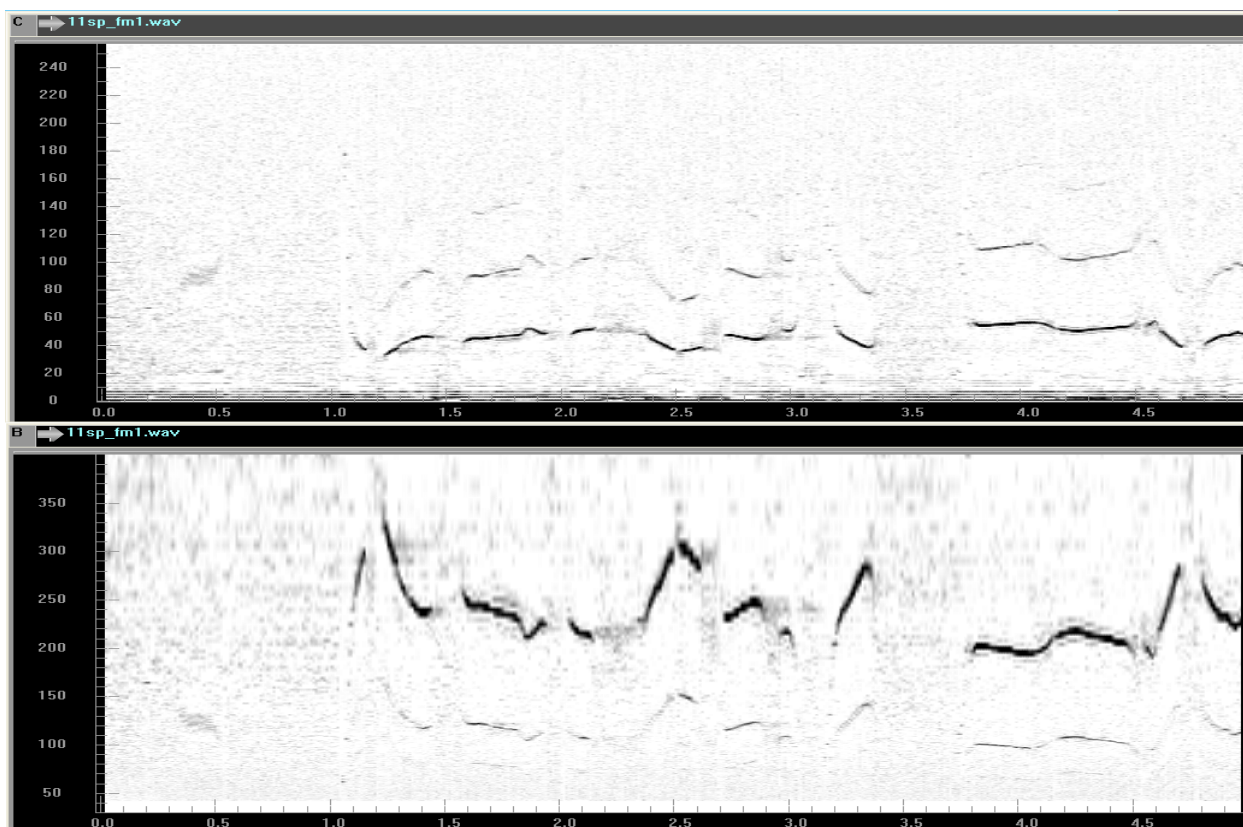


Рисунок 8 – Кепстрограмма речевого сигнала: наверху – временное представление (время-задержка-яркость), внизу - частотное представление (время-частота-яркость)

Применение кепстров [5]

Кепстры являются удобным инструментом для анализа периодических импульсных процессов (вибрации, тональные речевые сигналы), спектры которых характеризуются гребенчатой структурой. Шаг по частоте ΔF_0 между пиками гребенчатого спектра называется основным тоном и связан с периодом T_0 импульсного процесса соотношением:

$$\Delta F_0 = 1/T_0$$

При этом кепстр имеет максимум, соответствующий периоду T_0 .

Кепстрограмма является удобным инструментом для анализа процессов с периодической структурой спектра мощности (то есть со структурой основного тона, ОТ). Например, тональная речь, периодическая импульсная наводка, вибрации двигателей и пр. На кепстрограмме (в координатах время-частота) основной тон сигнала проявляется в виде темной линии на частоте ΔF_0 . На среднем кепстре (в координатах амплитуда кепстра – задержка) основной тон сигнала проявляется в виде максимума среднего кепстра для задержки T_0 .

Задание 1

Исследование белого и розового шума

Загрузите файл **16white_noise.wav** (белый шум). Постройте для сигнала кепстрограмму и далее вычислите средний кепстр.

Уменьшите амплитуду сигнала в e раз (2.73) и постройте средний кепстр. Как изменились коэффициенты среднего кепстра? Объясните.

Загрузите файл **16pink_noise.wav** (розовый шум). Постройте для сигнала кепстрограмму и далее вычислите средний кепстр.

В чем состоит качественное отличие средних кепстров розового и белого шума? С чем это связано?

Задание 2

Исследование периодического импульсного сигнала

Постройте для импульсного сигнала **11_xx_hz.wav** спектрограммы, на которых проявляются импульсные и волновые свойства сигнала (с малым и большим окном анализа).

Постройте средний спектр и попробуйте определить частоту основного тона сигнала по спектру.

Постройте осциллограмму и попробуйте определить частоту основного тона сигнала.

Постройте кепстрограмму и определите частоту основного тона сигнала.

Как период следования импульсов связан с частотой основного тона?

Задание 3

Исследование периодической структуры сигнала **16Landa_2012.wav**.

Определите период модуляции шума и частоту ОТ в сигнале по:

- огибающей мощности;
- осциллограмме;
- спектрограмме;
- среднему спектру;
- кепстрограмме.

Приведите соответствующие рисунки.

Задание 4

Тональная структура гласных РС мужчины и женщины

Постройте спектрограмму сигнала **11sp_fm1.wav** и найдите на ней участки ОТ мужчины и женщины.

Постройте кепстрограмму сигнала и определите интервалы частот ОТ мужчины и женщины.

Задание 5

Постройте кепстрограмму сигнала **00704bul.wav** и определите интервалы частоты ОТ и скорость её изменения (герц/сек).

Задание 6

Исследуйте тональную структуру и определите частоту ОТ сигналов

Писка комара — **komar.wav**;

Жужжания шмеля — **shmel.wav**;

Пения птицы — **bird.wav**.

Определите частоту ОТ издаваемых звуков.

ЛАБОРАТОРНАЯ РАБОТА № 6

Гистограмма и частота пересечения нуля

Цель работы: знакомство с оценками гистограммы и частоты пересечения нуля (Zero-crossing rate, ZCR) аудиосигналов и исследование свойств оценок этих характеристик.

Краткие теоретические сведения

Гистограмма [6]

Плотность вероятности является основной характеристикой случайных процессов. На экспериментальных данных можно получить выборочные вероятностные функции плотности или гистограммы.

Вычисление гистограммы производится следующим образом [3]. Выбирают (задают, определяют) интервал изменения величины $x(i)$: $a \leq x < b$, делят его на k подинтервалов равной длины («интервалов группировки») так, чтобы полная область изменения $x(i)$ распалась на $k + 2$ интервалов (включая интервалы $x < a, x \geq b$). После этого берут все данные (N) и фиксируют число (Nj) попаданий в каждый интервал. На полученных результатах можно построить последовательности трех видов:

Первая – последовательность $\{Nj\}$ без всяких изменений (гистограмма).

Вторая – последовательность $Pj = Nj/N$ – выборочных вероятностей, $j = 0, \dots, k + 1$. ($\sum Pj = 1$).

Третья – последовательность значений выборочной вероятностной функции, значения которой вычисляются по формуле:

$$Pj = \frac{Nj/N}{(b-a)/k} = Pj/\Delta, j = 0, 1, \dots, k + 1$$

Приведенной процедуре должен предшествовать выбор конкретных значений a, b , и k . Выбор обычно определяется предположением о распределении, которому подчиняются данные, способом сбора данных или предварительным анализом собранных данных. Оценки гистограммы зависят от задания наблюдаемого интервала изменения амплитуды процесса и размера подинтервалов (определяющих их количество на интервале наблюдения).

Частота пересечения нуля

Число пересечений нуля сигнала $x(n)$ на отсчет для прямоугольного окна размера N :

$$\begin{cases} \text{sgn}(x) = 1 & x \geq 0, \\ \dots = -1 & x < 0 \end{cases}$$

$$ZCR(k) = 1/2N \sum_{n=k-N+1}^k [\text{sgn}(x(n)) - \text{sgn}(x(n-1))]$$

где k – индекс кадра, n – индекс отсчетов сигнала.

Обычно задают временную длительность размер окна Tw . В этом случае размер окна составляет $N = Tw \times Fs$ точек.

Частота пересечения нуля является простой мерой частотных свойств сигналов, особенно узкополосных сигналов, например синусоид. Синусоида частоты Fo при частоте дискретизации Fs имеет Fs/Fo отсчетов на цикл и 2 пересечения нуля на цикл. Таким образом, синусоида имеет:

$ZCR = (2 \text{ пересечения} / \text{цикл}) \times (Fo/Fs \text{ циклов/отсчет}) = 2 Fo/Fs \text{ пересечений/отсчет} = 2 Fo \text{ пересечений} / \text{с}.$

На рисунке 9 приведены графики частоты пересечения нуля двух тестовых сигналов $Fs = 11025$ Гц – синусоиды с периодом 7 отсчетов ($Fo = 1575$ Гц) и белого шума $\sigma = 1000$.

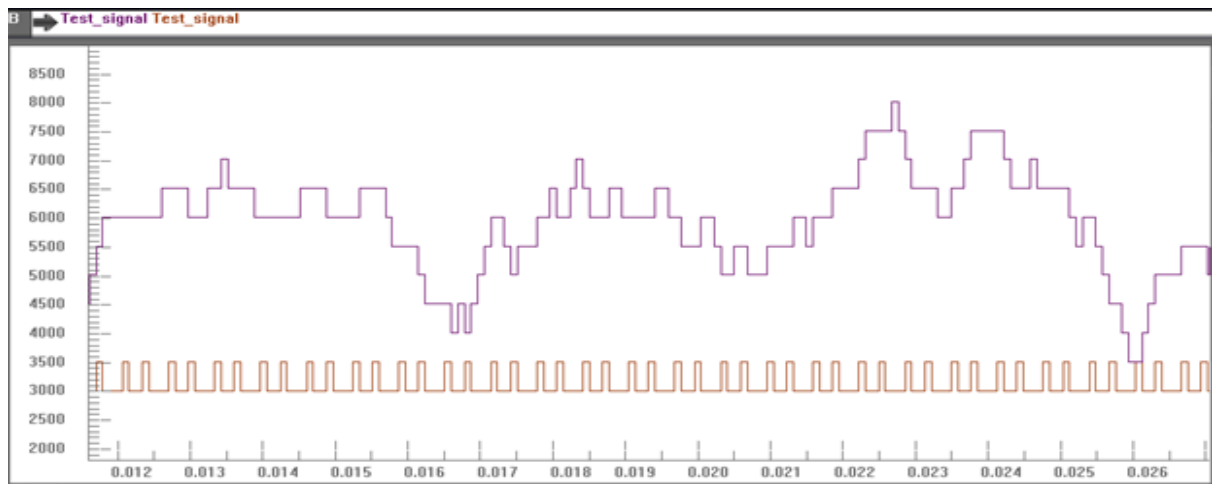


Рисунок 9 – Графики $ZCR(k)$ сигналов в окне 2 мс: нижний – синусоиды с периодом 7 отсчетов ($Fo = 1575$ Гц), верхний – белого шума $\sigma = 1000$

Задание 1

Возьмите сигнал белого шума **8k_white_noise.wav** и постройте в одном окне его гистограммы с шагом 10, 100, 1000.

Объясните, в чем заключается и чем объясняется различие оценок гистограмм?

Задание 2

Загрузите файлы *8kHz_sin.wav* и *8kHz_sin77.wav* постройте в одном окне их гистограммы с шагом 100. Объясните характер гистограмм.

Задание 3

Загрузите файл *SPCH_ME.wav* (речевой сигнал), постройте его гистограмму с шагом 100.

Возьмите речевой фрагмент сигнала и постройте его гистограмму с шагом 100.

Объясните, в чем заключается и с чем связано различие оценок гистограмм?

Задание 4

Загрузите файл *11_100HZ.wav*.

Постройте для него графики ZCR с разным размером окна анализа (размером кадра): 2 мс, 20 мс, 200 мс.

Как оценка ZCR связана с размером кадра?

Добавьте к сигналу константу (+2000) и постройте для него графики ZCR. Как изменится ZCR? Объясните результаты.

Задание 5

Постройте графики ZCR для чистого и зашумленного речевого сигнала (файлы *SPCH_ME.wav* и *SMN1_22.wav*).

В каких пределах изменяется ZCR?

На каких участках речевого сигнала ZCR увеличивается?

Добавьте к сигналу *SMN1_22.wav* константу (+2000). Как изменится ZCR? Объясните результаты.

Задание 6

Постройте графики ZCR для одного и того же речевого сигнала с разной частотой дискретизации (файлы *44sp_m.wav*, *22sp_m.wav*, *11sp_m.wav*).

Меняется ли ZCR при изменении частоты дискретизации (44 кГц, 22 кГц, 11 кГц) сигнала? Объясните результаты.

ЛАБОРАТОРНАЯ РАБОТА № 7

Отношение сигнал-шум и время реверберации

Цель работы: экспресс-оценка отношения сигнал/шум (ОСШ) речевых сигналов (РС) и времени реверберации на фонограммах по огибающим мощности сигналов.

Краткие теоретические сведения

Огибающие мощности сигналов – это последовательность значений средней величины квадрата амплитуды сигнала в скользящем окне заданной длительности T_w :

$$Px(t) = \langle x^2(t) \rangle / T_w = 1/N \sum_{i=t}^{t+N} x^2(i)$$

где $\langle \rangle$ – символ усреднения в окне, t – индекс времени, N – размер окна, $N = F_s \cdot T_w$.

Для центрированного процесса $Px(t) = \sigma_x(t)^2$.

Можно заметить, что оценки значения средней мощности в скользящем окне имеют близкие значения для разных размеров окон и разных частот дискретизации.

В программе SIS II под термином «Энергия» понимается корень мощности сигнала:

$$Ex(t) = \left[\frac{1}{N} \sum_{i=t}^{t+N} x^2(i) \right]^{1/2}$$

Для центрированного процесса $Ex(t) = \sigma_x(t)$.

Оценка отношения сигнал/шум

$$SNR \text{ дБ} = \langle 10 \log_{10} \left(\frac{Ps(t)}{Pn(t)} \right) \rangle = Ps(t) \text{ дБ} - Pn(t) \text{ дБ}$$

где $\langle \rangle$ – символ усреднения по речевым кадрам, $Ps(t)$, $Pn(t)$ – мощности (в программе SIS II вычисляются «энергии» $E(t)$, пропорциональные квадратному корню из мощностей) сигналов на кадрах.

В этом случае

$$SNR \text{ дБ} = \langle 20 \log_{10} \left(\frac{Es(t)}{En(t)} \right) \rangle = s$$

Поскольку речевой сигнал и шум являются независимыми сигналами, мощность зашумленного сигнала равна сумме мощностей шума и речи:

$$Px(t) \approx Ps(t) + Pn(t)$$

$$Ps(t) \approx Px(t) - Pn(t)$$

$$SNR \text{ дБ} \approx < 10 \log_{10} \frac{Px(t) - Pn(t)}{Pn(t)} >$$

В случае $Ps(t) \gg Pn(t)$ получаем на кадрах речи оценку мощности РС: $Ps(t) \approx Px(t)$, $Es(t) \approx Ex(t)$. В случае стационарного шума $Pn(t) \approx const$, средние оценки $En(t)$ можно сделать в паузах речи. Тогда

$$SNR \text{ дБ} \approx < 20 \log_{10}(Es(t)/En(t)) > = < Es(t) > \text{ дБ} - < En(t) > \text{ дБ}$$

Оценка времени реверберации

Время реверберации - время после выключения источника звука, за которое средняя плотность энергии звука $P(t)$ уменьшается в 10^6 раз, а «энергия» сигнала $E(t)$ уменьшается в 10^3 раз, то есть на 60 дБ.

После отключения источника звука огибающая мощности уменьшается по экспоненте, показатель которой тем больше, чем меньше время реверберации Tr .

Допустим, импульс начался в момент $t = 0$, когда величины огибающей мощности и «энергии» имели значения Po и Eo соответственно. Тогда в момент t в помещении с временем реверберации Tr характеристики $P(t)$ и $E(t)$ будут иметь следующие значения:

$$P(t) = Po \cdot 10^{-6 \cdot t/Tr} = Po \exp[-13,8 \cdot t/Tr]$$

$$E(t) = Eo \cdot 10^{-3 \cdot t/Tr} = Eo \exp[-6,9 \cdot t/Tr]$$

На графиках в логарифмическом масштабе (дБ) огибающая «энергии» уменьшается приблизительно линейно:

$$E(t) \text{ дБ} = Eo \text{ дБ} - (60 \cdot t/Tr) \text{ дБ}$$

Метод оценки времени реверберации по огибающей «энергии» состоит в следующем. Выберем на осциллограмме сигнала (графике $E(t)$ дБ) момент To начала затухания импульса. Найдем на графике огибающей импульса другой момент $To + T$, когда огибающая реверберационного отклика импульса в

логарифмическом масштабе достигает уровня шума и перестает уменьшаться линейно. Измерим уровни $E(T_0)$ и $E(T_0 + T)$:

$$E(T_0) = E_0$$

$$E(T_0 + T) = E(T_0) \cdot 10^{-3 \cdot T / Tr}$$

Оценим отношение величин $E(T_0)$ и $E(T_0 + T)$ в децибелах.

$$R \text{ дБ} = 20 \log \left[\frac{E(T_0)}{E(T_0 + T)} \right] = E(T_0)_{\text{дБ}} - E(T_0 + T)_{\text{дБ}} = 20 \cdot 3 T / Tr = 60 \cdot T / Tr$$

Тогда по величинам R дБ и T можно оценить время реверберации:

$$Tr = T \cdot 60 / R \text{ дБ}$$

Например, $R = 60$ дБ, $T = 1$ с. Тогда $Tr = 1$ с.

Например, $R = 15$ дБ, $T = 0,25$ с. Тогда $Tr = 1$ с.

Таким образом, для оценки времени реверберации необходимо определить уровень мощности («SIS-энергии») сигнала в начальный момент импульса. После этого найти время, за которое огибающая мощности уменьшается на 15, 20 или 30 дБ, и определить по нему время уменьшения на 60 дБ. Например, время уменьшения на 20 дБ равно T . Тогда $Tr = 3T$.

Задание 1

Подберите размер окна для анализа огибающей мощности тональных звуков мужской речи (*spch_me.wav*) и женской речи (*spch_we.wav*) и постройте огибающие мощности на фрагментах этих сигналов.

Оцените темп тональных звуков в секунду в РС *spch_me.wav*, *spch_we.wav*.

В каком динамическом диапазоне (дБ) находятся звуки РС?

Задание 2

Возьмите сигналы *SMN1_xx.wav*, постройте огибающие мощности («энергия», окно 20 мс) в логарифмическом масштабе и оцените ОСШ каждого из сигналов в децибелах.

Каким образом шум влияет на форму огибающих мощности («энергии»)?

Задание 3

Возьмите сигналы **14i_xx.wav** постройте огибающие мощности («энергия», окно 20 мс) в логарифмическом масштабе и оцените время реверберации каждого из сигналов в миллисекундах.

Задание 4

Возьмите сигналы **14i_xx.wav** постройте огибающие мощности в логарифмическом масштабе и оцените время реверберации каждого из сигналов в миллисекундах на нескольких (2-3) импульсах.

СПИСОК ЛИТЕРАТУРЫ

1 SIS II Специализированный звуковой редактор STC-S521. Руководство пользователя. ЦВАУ 00432-01 94.

2 Солонина А. И., Клионский Д. М., Меркучева Т. В., Перов С. Н. Цифровая обработка сигналов и MATLAB. Учебное пособие, СПб, БХВ-Петербург, 2013. 512 с.

3 Марпл С. Л. Цифровой спектральный анализ и его применение. М.: Мир, 1990. 584 с.

4 Маркел Дж. Д., Грэй А. Х. Линейное предсказание речи. М.: Связь, 1980. 308 с.

5 Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. М.: Мир, 1978. 841 с.

6 Отнес Р. Прикладной анализ временных рядов. М.: Мир, 1982. 428 с.

Часть 2. АЛГОРИТМЫ КОМПЕНСАЦИИ ИСКАЖЕНИЙ РЕЧЕВЫХ СИГНАЛОВ

ЛАБОРАТОРНАЯ РАБОТА № 1 Основы работы с программой Sound Cleaner II

Цель работы: ознакомление с основными функциональными возможностями программного комплекса Sound Cleaner II [1].

Краткие теоретические сведения

Искажения — изменение речевого сигнала, ухудшающее его качество.

Шум — наложенный на речь посторонний сигнал.

Факторы ухудшения качества речевых сигналов (РС):

- малая полоса пропускания канала связи;
- реверберация;
- высокий уровень шума окружения;
- акцентная и шепотная речь;
- тихая речь;
- непрерывная речь, голосовой коктейль, ломбард-эффект;
- нестационарный фоновый шум (музыка, уличный шум, шумы в транспорте);
- большая дистанция между диктором и микрофоном;
- нелинейные амплитудные искажения.

Основные источники искажений и шумов:

- шумы окружения (ветер, транспорт, компьютеры и др.);
- шумы в каналах связи;
- речь посторонних людей;
- шумы от аудиоустройств (телевизор, трансляция, радио и др.);
- электрические наводки (источников питания и др.);
- реверберация;
- переполнения и насыщение динамического диапазона;
- АЧХ канала.

Шумы и искажения уменьшают качество и разборчивость речевых сигналов и ухудшают автоматическое распознавание речи. Обработка РС с использованием алгоритмов шумоочистки позволяет частично компенсировать различные типы шумов и искажений.

Основные цели шумоочистки:

- восстановление параметров РС, прошедших через канал (Feature Enhancement);
- улучшение субъективного качества (восприятия) РС (Speech Quality Enhancement);
- увеличение разборчивости РС (Speech Intelligibility Enhancement).

Ознакомление с имеющимися фильтрами программного комплекса и их предназначением

Широкополосный фильтр: фильтр предназначен для снижения уровня шума помещений и улицы, помех каналов связи или записывающей аппаратуры. Фильтр является адаптивным и обладает максимальной эффективностью при простоте настроек.

Инверсный фильтр: фильтр адаптивной инверсной фильтрации предназначен для выравнивания амплитудно-частотной характеристики канала связи, в котором сделана аудиозапись. Настройки фильтра нацелены на приближение усредненного спектра к плоскому, что позволяет демаскировать речевой сигнал и повысить его разборчивость.

Фильтр тональных помех: фильтр позволяет удалять из сигнала как узкополосные стационарные помехи, так и помехи, имеющие регулярный характер, такие как вибрации, сетевые наводки, шумы бытовых приборов, медленную музыку.

Импульсный фильтр: адаптивная фильтрация импульсных помех предназначена для восстановления речевых или музыкальных сигналов, искаженных импульсными помехами, например, щелчками, помехами радиоэфира, шуршанием виниловой пластинки.

Динамический фильтр: фильтр динамической обработки предназначен для повышения разборчивости речи в условиях больших перепадов уровня сигнала.

Эквалайзер: 4096-полосный графический эквалайзер со встроенным спектрографом позволяет проводить детальную коррекцию спектра в искаженных аудиозаписях.

Дереверберационный фильтр: дереверберационный фильтр позволяет снизить уровень реверберации в аудиозаписях, с целью повышения разборчивости речи и снижения утомляемости при длительном прослушивании.

Подавитель DTMF сигналов: Фильтр предназначен для обработки сигналов тонального набора телефонного номера.

GSM-фильтр: фильтр позволяет ослаблять gsm-помехи.

Деклиппирование: алгоритм деклиппирования обнаруживает и восстанавливает перегруженные фрагменты аудиозаписи, реконструируя волновую форму на данных участках.

Автоматический фильтр: автоматический фильтр существенно понижает уровень наиболее распространенных типов шумов с использованием лишь одного вида управления.

Стереофильтрация: стереофильтрация строится на вычитании из основного канала шумов, присутствующих в опорном. Функция синхронизации каналов позволяет записывать звук опорного канала отдельно от основного.

Ознакомление с функциями редактирования

Нормализация сигнала.

Клиппирование сигнала.

Передискретизация сигнала.

Темпокоррекция.

Сtereo/mono операции.

Ознакомление с транскрайбером

Ознакомление с функцией «Формирование отчета»

Задание 1

Откройте файл *Наводка.wav*.

1. Воспроизведение

1.1. Масштабирование и прокрутка

Поместите курсор мыши на горизонтальную шкалу времени.

Используйте колесо прокрутки мыши для увеличения или уменьшения разрешения по оси времени.

Нажмите  (или **F8**), чтобы увидеть весь сигнал в области просмотра.

Нажмите  (или **F7**), чтобы автомасштабировать сигнал по вертикали.

Используйте область навигации над осциллограммой для быстрого перемещения по сигналу, удерживая выделенный интервал с помощью курсора мыши.

1.2. Воспроизведение

1.2.1. Воспроизведение всего файла

Нажмите  (или **Пробел**).

1.2.2. Воспроизведение от курсора

Нажмите левой кнопкой мыши на выбранном месте на осциллограмме.

Нажмите кнопку  (или **Пробел**, или **Воспроизведение** → **С позиции курсора**).

1.2.3. Воспроизведение выделенного фрагмента

Нажмите и удерживайте левую кнопку мыши от начала или конца выделенного фрагмента и двигайтесь вдоль сигнала вправо или влево, после чего отпустите кнопку.

Нажмите кнопку  (или **Пробел**, или **Воспроизведение** → **Выделенный фрагмент**).

1.3. Пауза

Нажмите кнопку  (или клавишу **Пробел**).

1.4. Остановить

Нажмите кнопку  (или **Esc.**)

1.5. Информация о файле

Информацию о файле можно посмотреть в строке состояния:

Либо: **Файл** → **Информация о файле**:

2. Редактирование

2.1. Редактирование / Удалить

Выделите фрагмент для удаления.

Нажмите клавишу **Delete** (или **Редактирование** → **Удалить**).

2.2. Редактирование / Обрезать

Выделите интересующий вас фрагмент.

Нажмите **Редактирование** → **Обрезать** (или **Ctrl+T**).

2.3. Редактирование / Разделить стерео на два моно

Откройте файл *Стереo.wav*.

Нажмите **Редактирование** → **Разделить стерео на два моно** (или **Ctrl+2**).

Выберите вкладку **Стереo_left**.

Нажмите **Файл** → **Экспорт в звуковой файл**.

3. Обработка

3.1. Обработка / Нормализация

Откройте файл *Нормализация.wav*.

Воспроизведите.

Нажмите **Обработка** → **Нормализация**.

Установите 25000 отсчетов. Нажмите кнопку **Применить**.

Воспроизведите.

Выделите фрагмент от 0,00 до 3,21 с.

Повторите предыдущие шаги.

Воспроизведите сигнал.

3.2. Обработка / Клиппирование

Откройте файл *Импульс.wav*.

Воспроизведите.

Оцените уровень полезного сигнала.

Нажмите **Обработка** → **Клиппирование**.

Установите 7000 отсчетов. Применить.

Воспроизведите.

3.3. Обработка / Передискретизация

Откройте файл *Радио.wav*.

Проверьте текущую частоту дискретизации.

Нажмите **Обработка** → **Передискретизация**.

Выберите «Делить на целое» и установите 4. Применить.

3.4. Обработка / Темпокоррекция

Откройте файл *Ралли.wav*.

Нажмите **Обработка** → **Темпокоррекция**.

Установите значение 0.6, чтобы замедлить темп речи на 30%. Нажмите кнопку **Применить**.

ЛАБОРАТОРНАЯ РАБОТА № 2

Импульсные помехи и динамическая обработка сигналов

Цель работы: получение навыков работы с основными функциональными возможностями импульсного и динамического фильтров.

Краткие теоретические сведения

Значительные перепады уровня сигнала приводят к уменьшению разборчивости РС вследствие временной маскировки сигналов. Основными причинами перепадов уровня сигнала являются импульсные помехи и различие уровней сигналов разных дикторов в канале. Для компенсации перепадов уровней сигналов применяют подавители импульсных помех и динамические фильтры.

Подавители импульсных помех

Существует большое число источников и типов импульсных помех.

Некоторые типы импульсных помех:

- периодические / случайные;
- электрические / механические;
- импульсы квантования (перебросы АЦП).

Алгоритмы компенсации импульсных помех включают в себя два основных шага: детектирование и интерполяцию/сглаживание импульсов. Детектирование импульсов в значительной степени определяется их характеристиками (длительность, область частот и пр.).

Простейшие детекторы импульсов строятся на основе отношения оценки мощности сигнала в точке к мощности окружения:

$$Q(t) = 2P(t)/[P(t-T) + P(t+T)] =$$
$$= 2 \sum_{t=M/3}^{2M/3} x^2(t) / \left\{ \sum_{t=1}^{M/3} x^2(t) + \sum_{t=2M/3}^M x^2(t) \right\}$$
$$Q(t) = \sum_{t=M/3}^{2M/3} x^2(t) / \text{Max} \left\{ \sum_{t=1}^{M/3} x^2(t), \sum_{t=2M/3}^M x^2(t) \right\}$$

После нормализации критерия детектирования

$$Q(t) \rightarrow Qn(t): 0 \leq Qn(t) \leq 1$$

Компенсация импульсов осуществляется путем замены значений сигнала на участках импульсов интерполированными значениями $e(t)$:

$$y(t) = [1 - Qn(t)]x(t) + Qn(t)e(t)$$

Пример компенсации импульса представлен на рисунке 10.

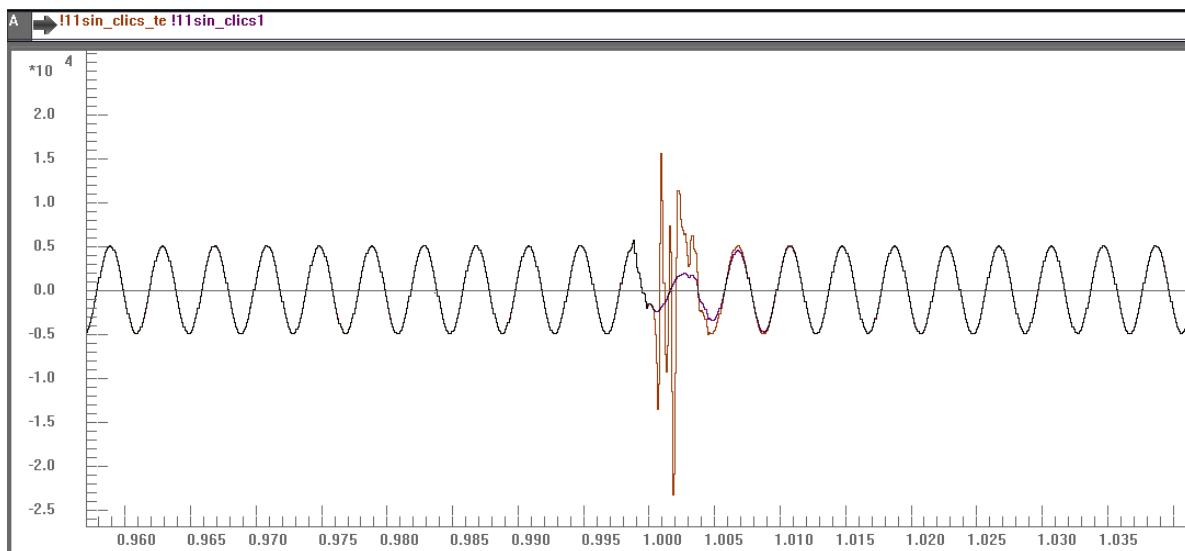


Рисунок 10 – Осциллограммы сигналов (временная шкала – сек): коричневый – исходный, серый – после обработки

Динамические фильтры

Основными амплитудными преобразованиями являются компрессор, лимитер, экспандер и гейт [1].

Компрессор – выравнивает уровни звуков: ослабление сильных, усиление слабых.

Лимитер (ограничитель) – пороговый компрессор с большим коэффициентом уменьшения динамического диапазона (ограничения громких звуков).

Экспандер (расширитель, пороговый шумоподавитель) – увеличивает динамический диапазон сигнала, понижая уровень громкости тихих сигналов.

Гейт (подавитель шума) – это пороговое устройство, отсекающее слабые сигналы, уровень которых находится ниже некоторого заданного порога.

Задачи, решаемые с помощью амплитудных преобразований:

- выравнивание уровня речи дикторов;
- приведение мощности сигнала к заданному уровню;
- подавление импульсов и ударов (защита слуха);
- управление коэффициентом усиления устройства звукозаписи;
- расширение динамического диапазона сигнала;
- подавление сигнала в паузах.

Динамические фильтры реализуются умножением входного сигнала на коэффициент передачи. Коэффициент передачи вычисляется не по мгновенным значениям входного сигнала, а по его мощности или средневыпрямленному значению. Поэтому преобразование в динамических фильтрах можно считать линейным.

Важными для слухового восприятия характеристиками является скорость ослабления сигнала в момент нарастания энергии речевого сигнала и скорость его усиления в момент спада энергии речевого сигнала.

Коэффициент передачи $G[t]$ определяется отношением средневыпрямленного или среднеквадратичного значения к значению (порогу), заданному пользователем.

Задание 1

Откройте файл *Импульс2.wav*.

Прослушайте запись.

Добавьте «Импульсный фильтр» в схему фильтрации.

Для лучшего результата установите «Степень подавления» 10 дБ.

Установите «Длительность импульса» 3,0 мс.

Добавьте «Широкополосный фильтр» и «Усилитель».

Отфильтруйте сигнал. Сохраните результат.

Задание 2

Откройте файл *dif.wav*.

Прослушайте запись.

Добавьте «Динамический фильтр».

Ослабьте сильный сигнал и усильте слабый сигнал для выравнивания разницы амплитуд.

Сохраните результат.

Задание 3

Откройте файл *Винил.wav*.

Осуществите самостоятельную обработку файла импульсным фильтром. Сохраните результат.

Откройте файл *Динам.wav*.

Осуществите самостоятельную обработку файла динамическим фильтром.

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 3

Частотная фильтрация

Цель работы: получение навыков работы с основными функциональными возможностями эквалайзера и частотной фильтрации.

Краткие теоретические сведения

Алгоритм частотной фильтрации

Представим линейную фильтрацию во временной области в виде свертки сигнала с импульсной характеристикой фильтра:

$$y(t) = h(t) * x(t),$$

где $h(t)$ – импульсная характеристика, $*$ – оператор свертки.

Представим фильтрацию кадра сигнала в спектральной области:

$X(n, k) = DFT\{X(t)\}$ – спектры входного кадра сигнала;

$Y(n, k) = DFT\{Y(t)\}$ – спектры выходного кадра сигнала;

$H(n) = DFT\{h(t)\}$ – передаточная функция (АЧХ).

В спектральной области фильтрацию можно представить в следующем виде

$$Y(n, k) = H(n)X(n, k) = Yr(n, k) + jYi(n, k)$$

где $Y(n, k)$ – спектр кадра сигнала на выходе фильтра,

Рассмотрим реальную и мнимую часть спектра выходного сигнала.

$$X(n, k) = Xr(n, k) + jXi(f, t)$$

$Xr(n, k)$ – симметричная, $Xi(n, k)$ – антисимметричная.

Поскольку $h(t)$ – вещественная и симметричная функция, то $H(n)$ – вещественная симметричная функция ($Hi(n) = 0$).

$$Y(n, k) = H(n)X(n, k) = HrXr + HrXi = Yr(n, k) + jYi(n, k)$$

Поскольку $Yr(n, k) = H(n)Xr(n, k)$ – симметричная функция, а $Yi(n, k) = H(n)Xi(n, k)$ – антисимметричная функция, то выходная последовательность на кадре $y(t) = IDFT\{Y(n, k)\}$ является вещественной функцией.

Эквалайзеры и фильтр темброкоррекции

Эквалайзеры – это управляемые спектральные фильтры. Назначение эквалайзеров:

- формирование сигнала с заданным спектром (spectral shaping);

- создание фильтров с частотной характеристикой сложной формы;
- коррекция спектральных искажений сигналов.

Типы эквалайзеров [2]:

- графические эквалайзеры (задаются полосы и коэффициенты передачи);
- параметрические эквалайзеры (задаются величины параметров передаточной функции);
- параграфический эквалайзер - это параметрический эквалайзер с графическим дисплеем;
- интеллектуальные эквалайзеры (с адаптивной подстройкой параметров).

Фильтр темброкоррекции (ФТК), в программе Sound Cleaner II – это 4-параметрический эквалайзер. Параметрами управления являются нижняя и верхняя частоты $F1, F2$ диапазона, где АЧХ равна единице и коэффициенты $K1, K2$ наклона (спада/подъема) АЧХ за пределами плоского участка АЧХ (в области низких и высоких частот).

На рисунке 11 приведен пример характеристики ФТК.

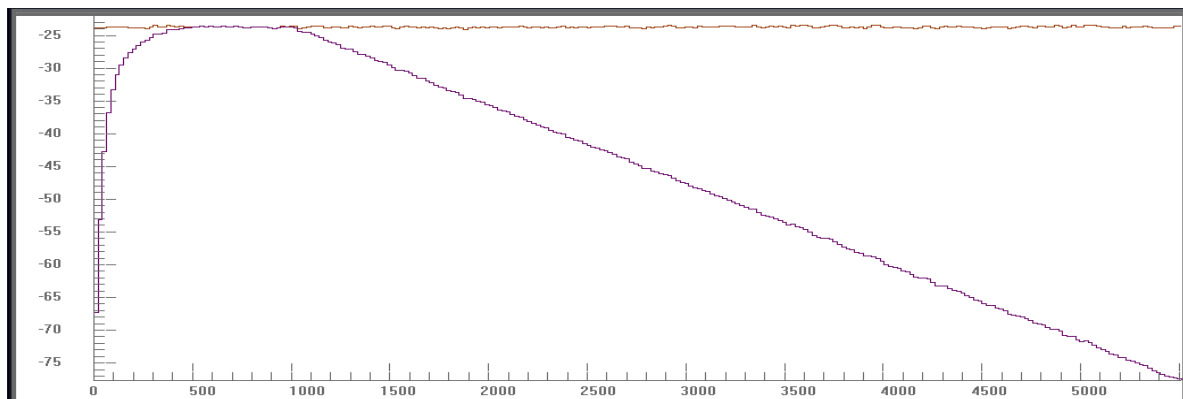


Рисунок 11 – Характеристика 4-параметрического эквалайзера
($F1=500$ Гц, $K1=-2$ дБ, $F2= 1000$ Гц, $K2 = -12$ дБ)

Задание 1

1) Откройте файл **Pub.wav**.

Воспроизведите запись.

Перенесите **Эквалайзер** из вкладки **Фильтры** в область левого канала справа (рис. 12).

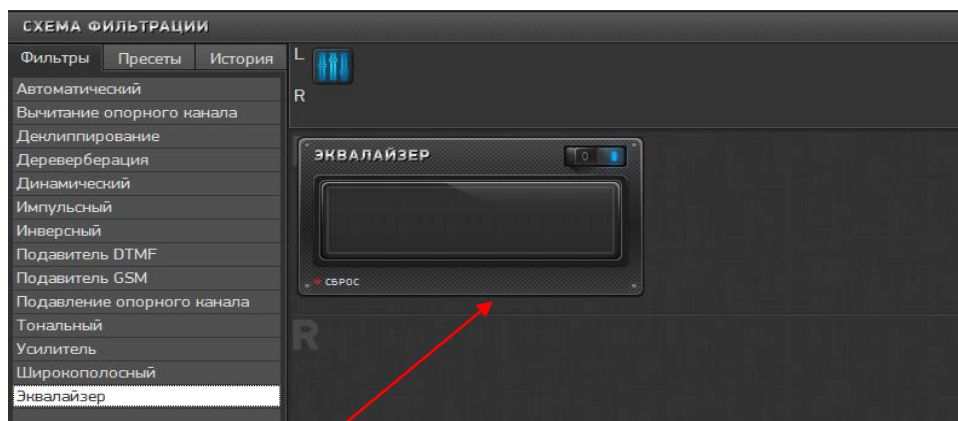


Рисунок 12 – Эквалайзер в интерфейсе ПО Sound Cleaner II

Двойным щелчком по экрану эквалайзера активируйте его настройки (рис. 13). Нажмите кнопку Σ , чтобы отобразить средний спектр исходного сигнала. Увеличьте разрешение по оси X в области низких частот с помощью колеса прокрутки мыши. Передвиньте слайдеры под частотами 60 Гц и 120 Гц вниз до -72 дБ.

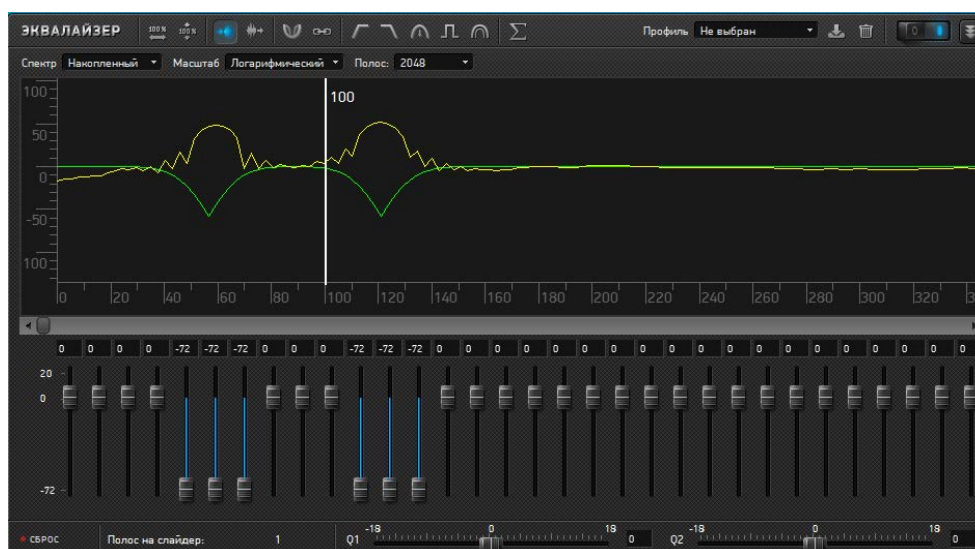


Рисунок 13 – Настройки эквалайзера (1)

Нажмите кнопку **Предпросмотр**  на панели управления воспроизведением.

Нажмите кнопку **Контрастировать фильтр** , чтобы усилить степень подавления помех.

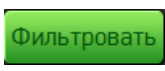
Нажмите кнопку **Спектр сигнала на выходе**  и переведите тип спектра на **Мгновенный**.

Воспроизведите запись.

Сравните спектры сигнала на входе и на выходе.

Нажмите кнопку , чтобы уменьшить размер окна **Эквалайзера**.

Добавьте «Усилитель» в схему фильтрации и передвиньте слайдер усилителя на 20 дБ.

Нажмите кнопку **Фильтровать** , чтобы применить фильтры.

Нажмите **Схема → Сохранить как...**, чтобы сохранить схему фильтрации и параметры фильтров в файл для дальнейшего использования.

Сохраните файл **Файл → Экспорт в звуковой файл... Pub_ф.wav**.

2) Откройте файл *Track09B.wav*.

Воспроизведите запись.

Откройте окно настроек **Эквалайзера** и нажмите кнопку , чтобы обнулить все слайдеры (рис. 14).

Оцените средний спектр сигнала на входе.

Используйте колесо прокрутки мыши, а также кнопки , для наилучшей визуализации спектра сигнала.



Рисунок 14 – Настройки эквалайзера (2)

Нажмите , чтобы построить инверсный фильтр (рис. 15, 16).

Передвиньте белую вертикальную марку справа на 4500 Гц, а левую марку на 199 Гц, чтобы изменить область обработки сигнала.



Рисунок 15 – Настройки эквалайзера (3)

Измените положение слайдера Q1 для темброкоррекции низких частот:
- 0 – для подавления низких частот и + 0 – для усиления низких частот.

Измените положение слайдера Q2 для темброкоррекции высоких частот:
- 0 – для подавления высоких частот и + 0 – для усиления высоких частот.



Рисунок 16 – Настройки темброкоррекции

Произведите фильтрацию и сохраните обработанный файл.

3) Откройте файл **Наводка.wav**.

Воспроизведите запись.

Оцените средний спектр сигнала.

Нажмите кнопку  **Гармонический фильтр**.

Произведите фильтрацию и сохраните обработанный файл как **Наводка_ЭК.wav**.

Задание 2

Откройте файл **Телефон.wav**.

Обработайте самостоятельно в соответствии с описанными ранее примерами.

ЛАБОРАТОРНАЯ РАБОТА № 4

Фильтрация широкополосного шума

Цель работы: получение навыков работы с основными функциональными возможностями фильтра широкополосного шума.

Краткие теоретические сведения

Алгоритмы фильтрации широкополосного шума

Источником широкополосного шума чаще всего является окружения: ветер, шумы транспортных средств, шумы компьютеров (вентиляторов), шумы бытовой техники (кондиционеры, холодильники) и др.

Для компенсации широкополосных шумов и выделения РС предложено большое число алгоритмов. Одним из основных является алгоритм спектрального вычитания, предложенный в работе [3].

Краткое описание алгоритма спектрального вычитания.

Модель сигнала:

$$x(t) = s(t) + w(t)$$

$$X(n, k) = S(n, k) + W(n, k).$$

$X(n, k), S(n, k), W(n, k)$ – дискретные спектры сигнала, речи, шума на кадрах, n – индекс частоты, k – индекс кадров

Фильтрация спектра сигнала:

$$Y(n, k) = G(n, k)X(n, k).$$

$Y(n, k)$ – спектр сигнала на выходе фильтра, $G(n, k)$ – АЧХ фильтра (вещественная целевая функция).

Целевая функция $G(n, k)$ вычисляется на основе следующих соотношений [3].

Поскольку РС и шум являются независимыми, то

$$Px(n, k) = Ps(n, k) + Pw(n, k)$$

$$Ps(n, k) = Px(n, k) - Pw(n, k) \approx Py(n, k) = Px(n, k) - |\hat{W}(n, k)|^2 = G^2(n, k)Px(n, k)$$

$$G(n, k) = [Px(n, k) - Pw(n, k)/Px(n, k)]^{1/2} = [SNR(n, k)/1 + SNR(n, k)]^{1/2} = [1 - Pw(n, k)/Px(n, k)]^{1/2}$$

$$SNR(n, k) = Ps(n, k)/Pw(n, k)$$

$Px(n, k), Ps(n, k), Pw(n, k)$ – дискретные спектры мощности сигналов $s(t), x(t), w(t)$ на кадрах, $|\widehat{W}(n, k)|^2$ – оценка спектра мощности шума $Pw(n, k)$ на кадре.

Оценка $|\hat{S}(n, k)|$ амплитудного спектра РС вычисляется на основе следующих соотношений. В качестве оценки мгновенного спектра шума на кадре берется оценка среднего спектра шума:

$$|\hat{S}(n, k)|^2 = |X(n, k)|^2 - |\widehat{W}(n, k)|^2 \approx |X(n, k)|^2 - \langle |W(n, k)|^2 \rangle$$

Тогда:

$$G(n, k) = [Px(n, k) - \langle |W(n, k)|^2 \rangle / Px(n, k)]^{1/2} = [1 - \langle |W(n, k)|^2 \rangle / Px(n, k)]^{1/2}$$

По оценке спектра мощности РС вычисляется оценка комплексного спектра РС:

$$\hat{S}(n, k) = |\hat{S}(n, k)| e^{j\Phi_x(n, k)}$$

$$e^{j\Phi_x(n, k)} = X(n, k) / |X(n, k)|$$

По оценке комплексного спектра РС с помощью обратного преобразования Фурье вычисляется оценка РС на кадре: $\hat{s}(t) = IDFT\{\hat{S}(n, k)\}$.

Другим вариантом метода спектрального вычитания является алгоритм амплитудного спектрального вычитания [3]. В основу алгоритма амплитудного спектрального вычитания положено следующее соотношение для амплитудных спектров речи и шума:

$$Ax(n, k) = As(n, k) + Aw(n, k)$$

$Ax(n, k), As(n, k), Aw(n, k)$ – дискретные амплитудные спектры сигналов $s(t), x(t), w(t)$ на кадрах.

Тогда целевая функция приобретает следующий вид:

$$G(n, k) = As(n, k) / Ax(n, k) = Ax(n, k) - Aw(n, k) / Ax(n, k) = 1 - \langle Aw(n, k) \rangle / Ax(n, k)$$

Физический смысл этих и других подобных алгоритмов заключается в подавлении спектральных компонент сигнала, амплитуды которых соизмеримы с амплитудой шума и сохранении остальных спектральных компонент.

Задание 1

Откройте файл *Street_noise.wav*.

Воспроизведите файл.

Добавьте «Широкополосный фильтр» в схему фильтрации.

Используйте слайдер «Глубины подавления» для установления интенсивности подавления широкополосного шума. Значение глубины подавления по умолчанию – 15 дБ.

Обратите внимание: чем интенсивнее степень подавления шума, тем менее естественным становится звучание речи. Рекомендуемые значения: от 8 дБ до 22 дБ.

Нажмите кнопку **Выбеливание**.

Данный режим фильтрации приводит к неравномерному по спектру подавлению шума, подавляя его на частотах, где его уровень больше.

Таким образом, данный режим приводит к «выбеливанию» спектра остаточного шума на выходе фильтра, что делает звучание профильтрованного сигнала более комфортным.

«Выбеливание» рекомендовано для всех типов широкополосных шумов, за исключением белого шума.

Задание 2

Откройте файл *Катер.wav*.

Осуществите самостоятельную обработку широкополосным фильтром.

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 5

Выделение речи на фоне тональных помех

Цель работы: получение навыков работы с основными функциональными возможностями фильтра тональных помех.

Краткие теоретические сведения

Существует большое число источников тональных помех (коррелированных помех в более широком смысле).

Тональные помехи (естественного происхождения):

- периодические импульсные процессы, порождающие в спектре гребёнку гармоник основного тона;
- сетевые наводки (50-60 Гц) и их гармоники;
- акустические шумы компьютера (пики спектра в НЧ области);
- звуки работающего двигателя;
- речь посторонних дикторов;
- голоса птиц.

Тональные сигналы искусственного происхождения:

- тональные импульсы в телефонной линии (DTMF и т. п.);
- сирены;
- музыка.

Для выделения РС, искаженных тональными помехами, предложено большое число алгоритмов (временных, спектральных и др.). Одним из наиболее эффективных является алгоритм компенсации тональных помех во временной области [4, 5].

Рассмотрим работу временного компенсатора тональных и коррелированных помех. На рисунке 17 представлена схема компенсатора.

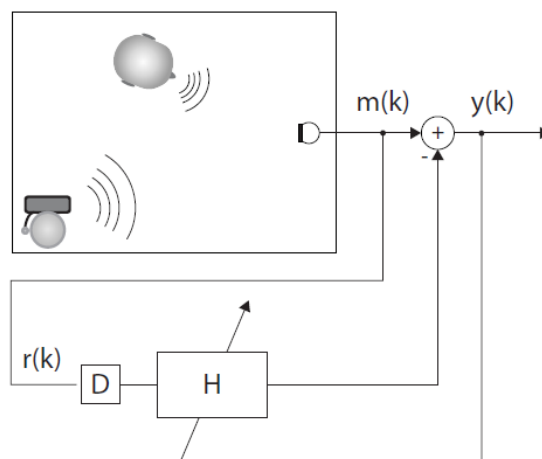


Рисунок 17 – Схема алгоритма компенсации коррелированных помех [4]

Модель сигнала микрофона

$$m(k) = s(k) + w(k)$$

Компенсация помехи осуществляется с помощью вычитания из зашумленного сигнала $m(k)$ оценки сигнала помехи $\hat{w}(k)$.

$$y(k) = m(k) - \hat{w}(k)$$

Оценка сигнала помехи вычисляется путем фильтрации задержанного на D отсчетов сигнала микрофона (интервал предсказания). Фильтр предназначен для предсказания величины помехи в момент k по задержанным значениям сигнала

$$R(k) = [m(k - D), m(k - D - 1), \dots, m(k - D - n + 1)]^T$$

$$\hat{w}(k) = r(k) * \hat{h}(k) = H^T R(k),$$

где $H = [h_1, h_2, \dots, h_n]^T$, $*$ – символ свертки.

Предсказание сигнала (на десятки и сотни отсчетов) возможно в пределах радиуса корреляции помехи.

Коэффициенты фильтра-предсказателя вычисляются на основе минимизации СКО ошибки (MSE) между сигналом микрофона $m(k)$ и выходом фильтра:

$$\langle |y(k)|^2 \rangle = \langle |m(k) - r(k) * h(k)|^2 \rangle$$

где $\langle \rangle$ – символ усреднения по времени.

Минимизация СКО ($\text{Min MSE} = \text{MMSE}$) приводит к оптимальному решению:

$$CrrHopt = Cmr \Rightarrow Hopt = Crr^{-1}Cmr$$

где Crr – ковариационная матрица сигнала $R(k)$, Cmr – вектор кросс-корреляции вектора $R(k)$ и сигнала $m(k)$.

Обращение матрицы (для фильтра с несколькими сотнями коэффициентов) требует значительных вычислительных затрат. Поэтому для оценки коэффициентов фильтра обыкновенно используется различные варианты адаптивного алгоритма [5]:

$$H(k) = H(k - 1) + \mu(k)e(k)R(k) = \mu(k)[m(k) - H^T(k - 1)R(k)]R(k)$$

где $\mu(k)$ – параметр адаптации.

Задание 1

Откройте файл *Pub.wav*.

Прослушайте запись.

Добавьте фильтр тональных помех.

Выберите **Помехи сети питания** или **Размер окна анализа 8192** отсчета.

Осуществите фильтрацию.

Сохраните результат.

Задание 2

Откройте файл *Телефон.wav*.

Осуществите самостоятельную обработку файла тональным фильтром.

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 6

Деклиппирование и дереверберация

Цель работы: получение навыков работы с основными функциональными возможностями фильтра дереверберации и фильтра деклиппирования.

Краткие теоретические сведения

Фильтр деклиппирования

В момент записи при перегрузке сигнала (выхода амплитуд за пределы динамического диапазона устройства) возникает ограничение амплитуды сигнала или клиппирование.

Клиппирование речевых сигналов проявляется, как правило, на громких тональных звуках и сопровождается «размножением» тональных компонент спектра в область высоких частот.

На рисунках 18, 19 приведены спектрограммы и спектры исходного неискаженного сигнала и сигнала после клиппирования.

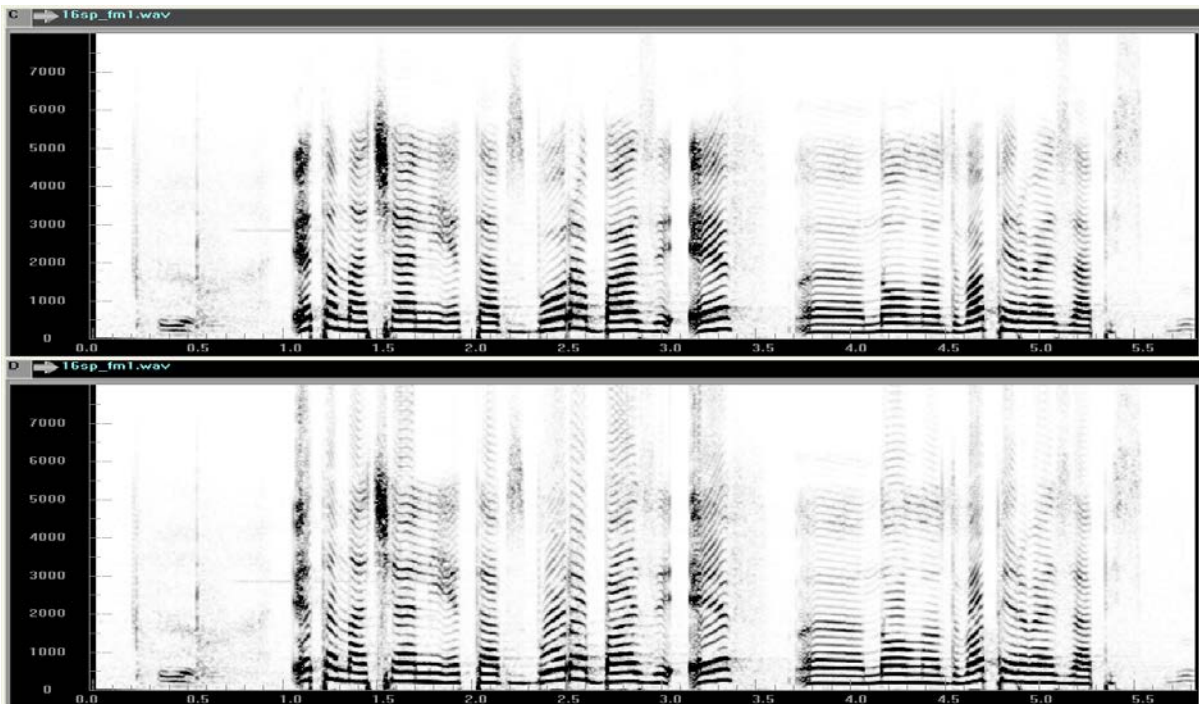


Рисунок 18 – Спектрограммы сигнала: верхняя исходного, нижняя – клиппированного

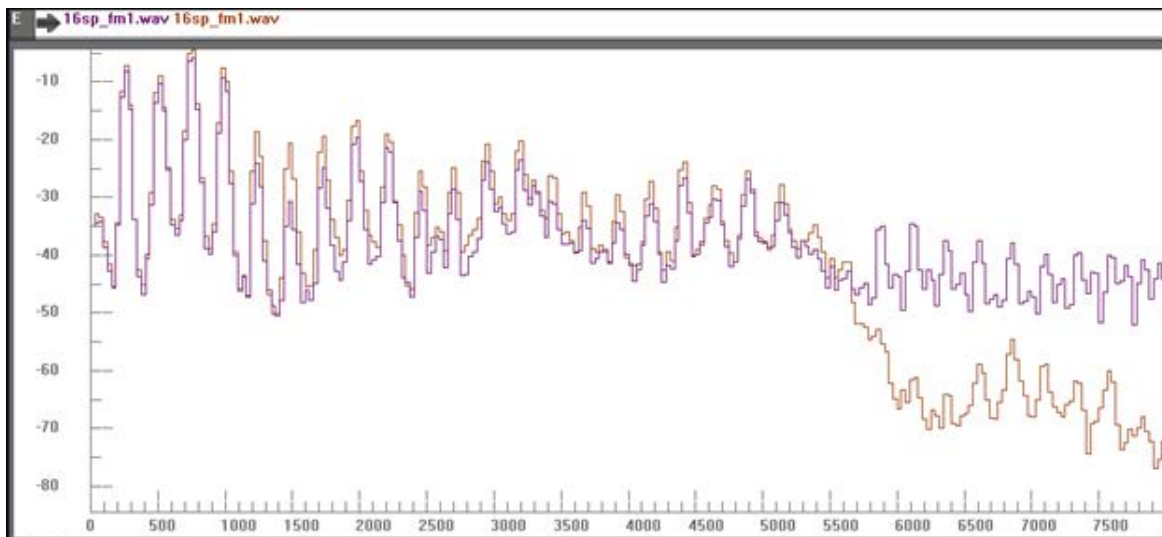


Рисунок 19 – Мгновенные спектры фрагмента исходного и клиппированного сигнала

Алгоритмы восстановления перегруженных фрагментов можно разделить на две группы: алгоритмы подавления участков клиппирования и алгоритмы интерполяции сигнала на участках клиппирования.

Алгоритм подавления клиппирования состоит в следующем.

По текущему фрагменту сигнала оценивается индикатор клиппирования $R(t) \gg 1$ в моменты клиппирования, $R(t) < 1$ в отсутствии клиппирования.

Далее рассчитывается нормализованный критерий клиппирования:

$$G(t) = R(t)/(1 + R(t)) \in [0 \dots 1]$$

Сигнал разделяется на низкочастотную $LF\{X(t)\}$ и высокочастотную $HF\{X(t)\}$ компоненты

$$X(t) = Xl(t) + Xh(t) = LF\{X(t)\} + HF\{X(t)\} = LF\{X(t)\} + (1 - LF)\{X(t)\}$$

Далее на участках клиппирования подавляется высокочастотная часть.

$$Y(t) = Xl(t) + (1 - G(t))Xh(t)$$

Фильтр дереверберации

Ревверберация — звук, образующийся в замкнутом пространстве в результате многократных отражений от отражающих поверхностей (диффузное поле). Ревверберация — ряд задержанных копий сигнала, приходящих к слушателю или микрофону из различных направлений.

В результате реверберации РС изменяется их спектр и появляется дополнительный шум. Таким образом, реверберацию РС можно рассматривать

как нестационарный широкополосный шум, порождаемый самим речевым сигналом. Ввиду того, что реверберационный шум возникает в результате множества отражений, можно считать его не коррелированным с речевым сигналом.

Поэтому одним из подходов к дереверберации является подавление этого нестационарного шума с помощью фильтров подавления широкополосного шума на основе метода спектрального вычитания. Главной особенностью алгоритма дереверберации является оценка спектра нестационарного шума.

Основываясь на предположении о том, что спектр мощности (или амплитудный спектр) шума пропорционален спектру мощности в предыдущий момент времени, можно оценить амплитудный спектр чистого сигнала следующим образом [6]:

$$|Y(f, k)| = |X(f, k)| - \exp(-6.9 Td/Tr) < |X(f, k - Td)| >$$

где Td – временной сдвиг между текущим и опорным кадром, Tr – время реверберации, $< >$ - оператор усреднения спектров по времени.

Тогда целевую функцию фильтра дереверберации можно записать в следующем виде:

$$\begin{aligned} G(f, k) &= (|X(f, k)| - \exp(-6.9 Td/Tr) < |X(f, k - Td)| >) / |X(f, k)| = \\ &= 1 - C \times < |X(f, k - Td)| > / |X(f, k)| \end{aligned}$$

Тогда выходной сигнал на кадре оценивается следующим образом

$$Y(t, k) = IDFT\{Y(f, k)\} = IDFT\{G(f, k)X(f, k)\}$$

Задание 1

Откройте файл **Clipped.wav**.

Прослушайте файл.

Добавьте **Деклиппирование** в схему фильтрации.

Отрегулируйте «Глубину подавления» для достижения наилучшего результата.

Для удаления высокочастотных щелчков используйте НЧ-фильтр. Он подавляет частотные компоненты сигнала расположенные выше 3 кГц.

Добавьте «Эквалайзеры» в левый и правый каналы и увеличьте значение Q2 для усиления высоких частот для более комфортного прослушивания.

Сохраните результат.

Задание 2

Откройте файл *reverb.wav*.

Прослушайте файл.

Добавьте фильтр **Дереверберация**.

Отрегулируйте параметр «Глубина подавления» для достижения наилучшего результата.

Сохраните результат.

Задание 3

Откройте файл *Клип.wav*.

Осуществите самостоятельную обработку файла фильтром «Деклиппирование».

Сохраните результат.

Откройте файл *Ревер.wav*.

Осуществите самостоятельную обработку файла фильтром «Дереверберация».

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 7

Выравнивание спектра

Цель работы: получение навыков работы с основными функциональными возможностями инверсного фильтра.

Краткие теоретические сведения

Фильтр адаптивной инверсной фильтрации предназначен для выравнивания амплитудно-частотной характеристики канала связи, в котором сделана аудиозапись. Настройки фильтра нацелены на приближение усредненного спектра к плоскому, что позволяет демаскировать речевой сигнал и повысить его разборчивость.

Прохождение РС через различные каналы (пространственные, электрические, радио) нередко приводит к искажению спектра РС. Искажение спектра способно привести к ухудшению качества РС (вследствие спектральной маскировки) и ухудшению его автоматического распознавания (вследствие изменения параметров сигнала относительно использованных на этапе обучения). Для компенсации такого рода эффектов нередко используют так называемые алгоритмы инверсной фильтрации.

Одним из наиболее широко используемых алгоритмов является алгоритм спектральной инверсной фильтрации (СИФ). Алгоритм СИФ предназначен для выравнивания среднего спектра сигнала. Иллюстрация этой процедуры приведена на рисунке 20.

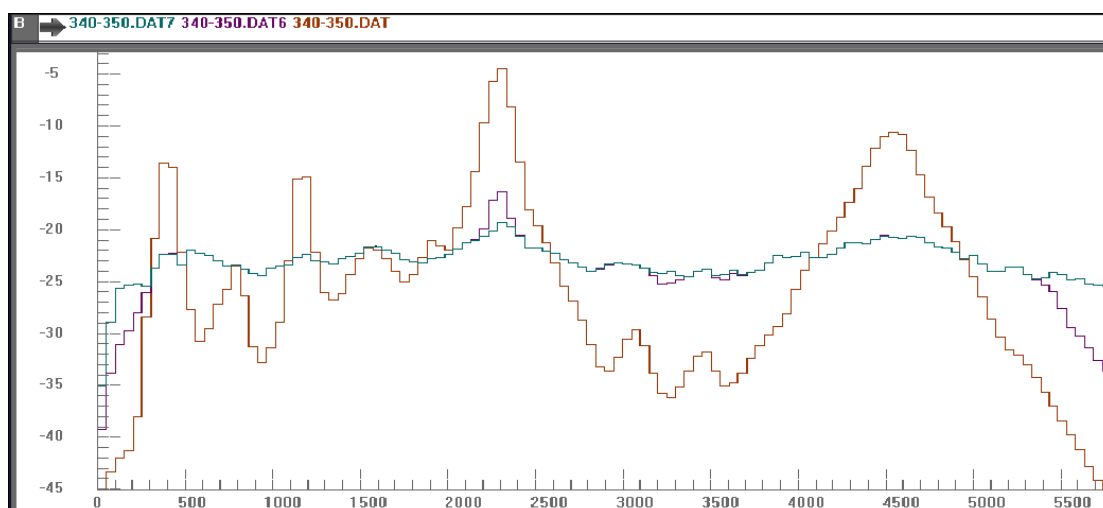


Рисунок 20 – Средний спектр сигналов: коричневый – исходного, зеленый – после СИФ

Алгоритм СИФ заключается в следующем.

Вектор отсчетов входного сигнала $X(t) = [x(t), x(t-1), \dots, x(t-N+1)]^T$ преобразуется на каждом кадре в спектр:

$$X(n, k) = DFT\{X(t)\}$$

где n, k – индексы частоты и кадра, соответственно.

На последовательности кадров вычисляется текущая оценка среднего амплитудного спектра (long-term spectrum):

$$\langle |X(n, k)| \rangle = \langle |X(n, k-1)| \rangle + \alpha[|X(n, k)| - \langle |X(n, k-1)| \rangle]$$

где α – коэффициент сглаживания спектра, $\langle \rangle$ – символ усреднения по времени.

Выравнивание спектра означает приведение среднего спектра сигнала к константе с помощью инверсного фильтра $H(n, k)$:

$$\text{Min} [H(n, k) \langle |X(n, k)| \rangle - C]^2 \Rightarrow H(n, k) = C / \langle |X(n, k)| \rangle$$

С помощью инверсного фильтра спектры входного сигнала преобразуются:

$$Y(n, k) = H(n, k)X(n, k)$$

и вычисляется кадр выходного сигнала

$$Y(t, k) = IDFT\{Y(n, k)\}.$$

Константа C определяется, например, исходя из условия равенства средних мощностей входного и выходного сигналов:

$$P_y(k) \approx P_x(k) \Rightarrow C(k)$$

СИФ может применяться для подавления помех, спектр которых имеет ярко выраженные максимумы. Кроме того, СИФ может использоваться в задачах автоматического распознавания для компенсации влияния АЧХ канала. В этом качестве СИФ эквивалентна процедуре вычитания среднего кепстра (Spectral Mean Subtraction, CMS) [7]:

$$|\hat{S}(k)|^2 = \exp[DFT\{Cx(k) - \langle Cx \rangle\}] = |X(k)|^2 / \langle |X(k)|^2 \rangle$$

где $\langle \rangle$ – символ усреднения по времени, $Cx(k)$ – кепстр сигнала на кадре k .

Задание 1

Откройте файл *Inverse.wav*.

Воспроизведите файл.

Добавьте «**Инверсный фильтр**» в схему фильтрации.

Для настройки работы инверсного фильтра используйте следующие параметры:

1. Подавление спектральных компонент устанавливает максимальное подавление сильных спектральных компонент сигнала. Подавление необходимо, чтобы привести неравномерный спектр шума к плоскому (белому) спектру, ослабив его максимумы. Рекомендуемые значения – 12-24 дБ.

2. Усиление спектральных компонент устанавливает предельное усиление слабых компонент спектра сигнала. Данный параметр регулирует степень выравнивания спектра за счёт усиления слабых компонент (обычно в области высоких частот). Ограничение максимального усиления необходимо, чтобы не усиливать широкополосный шум в области высоких частот. Рекомендуемые значения – 12-24 дБ.

Кнопка **Речевой эквалайзер** включает режим приведения среднего спектра сигнала к модельному среднему спектру речевого сигнала (рис. 21). В ряде случаев сигнал с таким спектром звучит комфортнее сигнала с плоским средним спектром

Разместите **Эквалайзер** до и после **Инверсного фильтра**, чтобы сравнить спектр сигнала на входе и выходе.

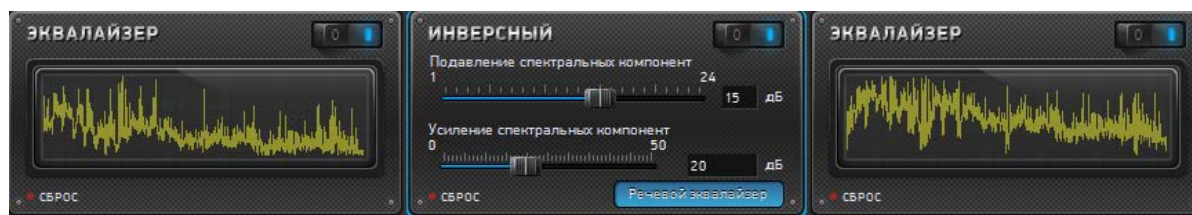


Рисунок 21 – Схема фильтрации

Откройте файл **AVTST.wav**.

Прослушайте запись.

Активируйте «Инверсный фильтр».

Добавьте «Широкополосный фильтр».

Осуществите обработку. Сохраните результат.

Задание 2:

Откройте файл **Радио.wav**.

Осуществите самостоятельную обработку инверсным фильтром.

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 8

Подавление помех, вызванных работой аппаратов связи

Цель работы: получение навыков работы с основными функциональными возможностями GSM-фильтра и подавителя DTMF сигналов.

Краткие теоретические сведения

GSM-наводки и DTMF-сигналы в телефонной линии представляют собой сигналы со сложной структурой спектра, частотный диапазон которого пересекается с частотным диапазоном речи. Это влечет за собой ухудшение разборчивости речи. Ослабление GSM-наводки и DTMF-сигналов является важной задачей.

Примеры спектрограмм речи на фоне GSM-наводки и DTMF-сигналов приведены на рисунках 22 и 23.

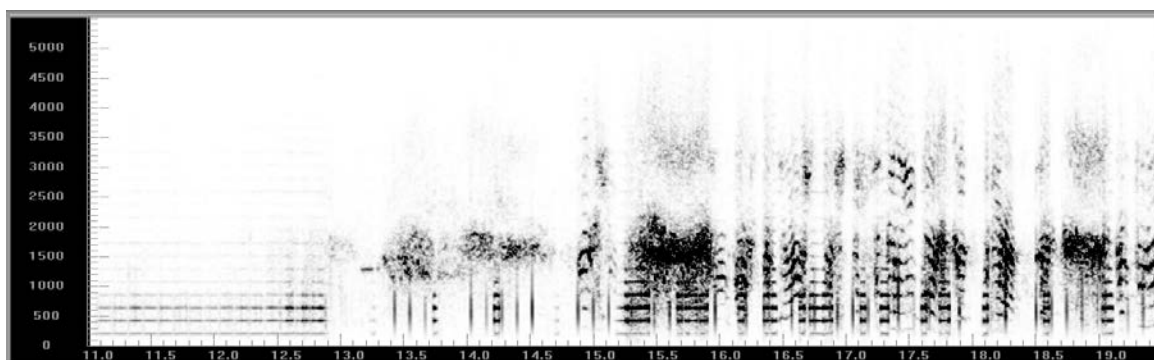


Рисунок 22 – Спектрограмма речи на фоне GSM-наводки

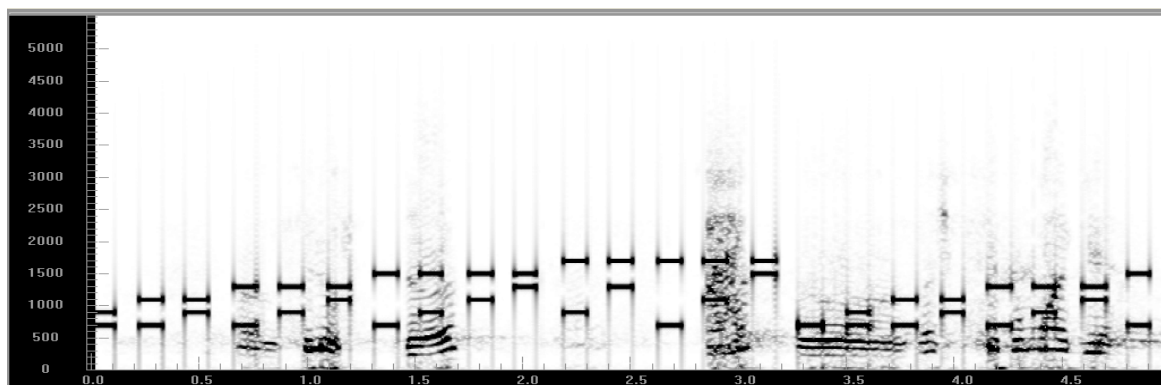


Рисунок 23 – Спектрограмма речи на фоне DTMF-сигналов

Подавитель GSM-наводки [4]

GSM-наводка – это периодические импульсы (120 мс) в полосе до 1500 Гц (с максимумом спектра в области 600 Гц) и периодические импульсы (4,615 мс) с частотой основного тона $F_0 = 217$ Гц. Тональные импульсы могут иногда пропадать, в этом случае остаются просто импульсы. На рисунках 24 и 25 представлена временная и спектральная структура GSM наводки.

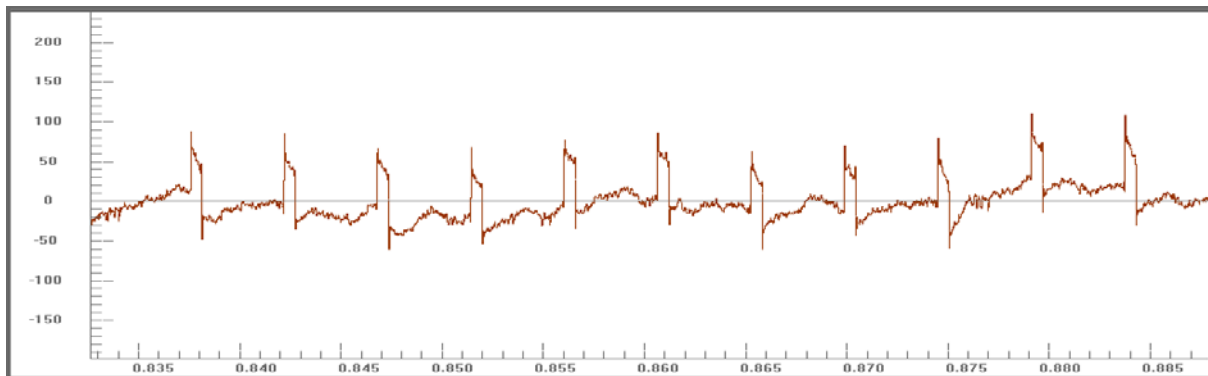


Рисунок 24 – Временная структура GSM-наводки

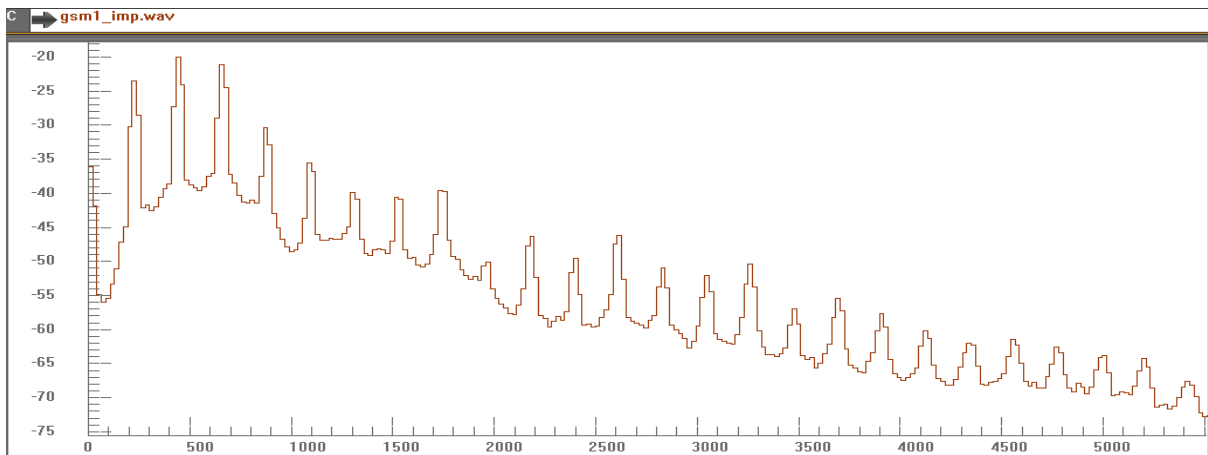


Рисунок 25 – Спектр GSM-наводки

Одним из способов подавления GSM наводки является обнаружение и подавление импульсов по образцу. Поскольку частоты $\{F_i = F_o \times i\}$ периодической импульсной наводки известны, то моменты появления наводки могут детектироваться по увеличению мощности сигнала в соответствующих полосах:

$$Q(t) = \sum_{i=1}^K P[F_i, t] / P(t) > THR$$

где $P[F_i, t]$ – текущая мощность сигнала в полосе F_i , $P(t)$ – текущая мощность сигнала.

Подавление GSM-наводки может осуществляться с применением последовательности 1-полосных фильтров-выкалывателей (notch-фильтров):

$$y(t) = \prod_{i=1}^K \text{Notch}(F_i)\{x(t)\}$$

Адаптивный алгоритм подавление сигналов DTMF

DTMF сигналы представляют собой тональные импульсы, в которых комбинация верхней и нижней частот $\{FL_i, FH_j\}$, $i = 1, \dots, 4$, $j = 1, \dots, 4$ кодируют символы.

Алгоритм подавления основывается на обнаружении тональных импульсов и подавления сигнала на соответствующих частотах в моменты детектирования импульсов. Обнаружение DTMF сигнала осуществляется с помощью оценки мощностей сигналов на выходе щелевых фильтров:

$$Q(iL, jH) = \{PL[iL]/\sum PL[i] > THR\} \wedge \{PH[jH]/\sum PH[j] > THR\}$$

где $PL[iL]$ – текущие мощности сигнала в полосах нижних частот DTMF-сигнала, $PH[jH]$ – текущие мощности сигнала в полосах верхних частот DTMF-сигнала.

Детектирование символов:

$$iL = \text{Argmax}\{PL[i]/\sum PL[i]\}$$

$$jH = \text{Argmax}\{PH[j]/\sum PH[j]\}$$

Подавление DTMF-сигнала осуществляется с применением последовательности двух фильтров-выкалывателей $\text{Notch}(FL_i) \text{Notch}(FH_j)$:

$$y(t) = \text{Notch}(FL_i)\text{Notch}(FH_j)\{x(t)\}$$

где $\text{Notch}(FL_i)$ – фильтр-выкалыватель на частоте FL_i , $\text{Notch}(FH_j)$ – фильтр-выкалыватель на частоте FH_j

Задание 1

Откройте файл **GSM.wav**.

Прослушайте запись.

Добавьте «**Эквалайзер**» и оцените средний спектр.

Добавьте «**Подавитель GSM**» с настройками по умолчанию.

Отрегулируйте «**Глубину подавления**» для достижения наилучшего результата.

Если включён «**Автодетектор**», то производится автоматический анализ обрабатываемого сигнала на наличие GSM-помехи. При выключенном детекторе фильтр-подавитель включён постоянно.

При управлении детектором используйте следующие рекомендации:

При сильной помехе и слабом полезном сигнале – включать детектор;

При слабой помехе и сильном полезном сигнале – пробовать оба режима.

При работе подавителя в начале и в конце GSM-помехи могут возникать короткие импульсы, вызванные переходными процессами в фильтре-подавителе. На осциллограмме данные импульсы выглядят как всплески сигнала, соответствующие началу и концу помехи, а на слух воспринимаются как периодически повторяющиеся щелчки. Для устранения подобных импульсов рекомендуется использовать «Подавитель щелчков».

Задание 2

Откройте файл *DTMF.wav*.

Прослушайте запись.

Добавьте «**Подавитель DTMF**».

Отрегулируйте «Глубину подавления» для достижения наилучшего результата.

Задание 3

Откройте файл *ДТМФ.wav*.

Осуществите самостоятельную обработку файла фильтром «**Подавитель DTMF**».

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 9

Выделение речи с использованием алгоритмов стереофильтрации

Цель работы: получение навыков работы с основными функциональными возможностями стереофильтра.

Краткие теоретические сведения

Алгоритм компенсации помехи опорного канала

Схема адаптивного компенсатора помехи опорного канала [4] (Adaptive Noise Canceller, ANC) приведена на рисунке 17.

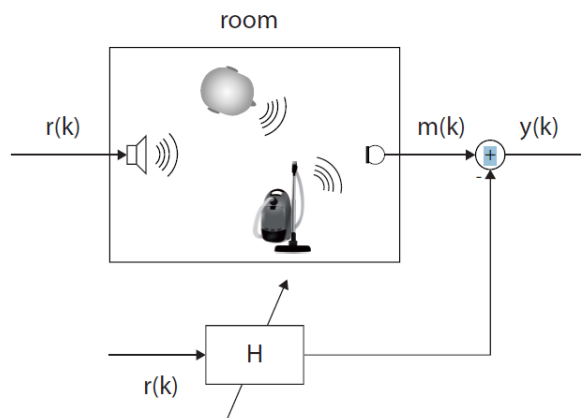


Рисунок 26 – Схема компенсации помехи

Модель сигналов:

$r(k) = a_{w1} * w(k)$ – сигнал опорного канала;

$m(k) = s(k) + a_{w2} * w(k)$ – сигнал основного канала.

где a_{w1}, a_{w2} – передаточные функции от источника помехи к опорному и основному каналу; $*$ – символ свертки; $s(k), w(k)$ – полезный сигнал и помеха.

Оценка полезного сигнала $\hat{s}(k)$ основана на минимизации мощности разности между сигналом основного канала и выходом фильтра помехи опорного канала:

$$\hat{s}(k) = y(k) = m(k) - r(k) * h(k)$$

Модель компенсации помехи в матричной форме:

$$\hat{s}(k) = y(k) = m(k) - H^T R(k)$$

где $R^T(k) = [r(k), r(k-1), \dots, r(k-N+1)]$ – вектор отсчетов сигнала опорного канала; $H^T = [h_1, h_2, \dots, h_n]$ – вектор весовых коэффициентов фильтра.

Минимизация СКО разности (ошибки) $y(k)$ между сигналом микрофона $m(k)$ и выходом фильтра $H^T R(k)$

$$\begin{aligned} \langle |y(k)|^2 \rangle &= \langle |m(k) - H^T R(k)|^2 \rangle = \langle |m(k) - H^T R(k)|^2 \rangle = \\ &= P_m - 2H^T C_{mr} + H^T C_{rr} H \end{aligned}$$

приводит к оптимальному решению минимума СКО (MMSE, Min MSE):

$$C_{rr} H_{opt} = C_{mr} \Rightarrow H_{opt} = C_{rr}^{-1} C_{mr} = C_{mr} / C_{rr}$$

где $\langle \rangle$ - операции усреднения по времени,

$C_{rr} = \langle R(k) R^T(k) \rangle$ - корреляционная матрица сигнала опорного канала,

$C_{mr} = \langle R(k) m(k) \rangle$ - вектор кросс-корреляции сигналов основного и опорного каналов.

Поскольку акустическая обстановка меняется, необходимо подстраивать коэффициенты фильтра. Однако вычисление обратной матрицы C_{rr}^{-1} требует значительных вычислительных затрат. Поэтому широкое распространение получило семейство алгоритмов, в которых коэффициенты фильтра вычисляются на основе последовательной минимизации критерия СКО (Least Mean Square, LMS) без обращения матрицы. Эти алгоритмы называются LMS-алгоритмами [4, 5].

Оценка коэффициентов фильтра осуществляется следующим образом. После прихода очередного отсчета на основной и опорный вход компенсатора вычисляется значение сигнала на выходе компенсатора:

$$y(k) = m(k) - H^T(k-1)R(k)$$

Коэффициенты фильтра вычисляются на основе следующего рекурсивного уравнения:

$$H(k) = H(k-1) - \mu(k) \times y(k) R(k)$$

где $\mu(k)$ – параметр скорости сходимости (в LMS-алгоритме).

Алгоритм подавления помехи опорного канала

Алгоритм компенсации помехи работает до тех пор, пока помеха в основном и опорном каналах является коррелированной. В замкнутых помещениях с увеличением расстояния между точками приема сигналов опорного и основного каналов корреляция сигналов помехи уменьшается вследствие реверберации [4]. В этом случае могут быть применены алгоритмы подавления

помехи опорного канала [7-9]. Хотя когерентность (фазовая корреляция) между сигналами каналов отсутствует, тем не менее, временные огибающие спектров сигналов остаются взаимосвязанными.

Алгоритм подавления помехи опорного канала является реализацией алгоритма спектрального вычитания, в котором оценка спектра шума вычисляется по сигналу опорного канала. Рассмотрим один из вариантов данного алгоритма [8, 9]. Запишем спектры сигналов основного и опорного каналов:

$$X(f, k) = S(f, k) + W(f, k)$$

$$R(f, k) = Wr(f, k)$$

где $W(f, k)$, $Wr(f, k)$ – спектры шума в основном и опорном каналах, k – индекс кадра.

Шум в основном канале является реверберационным откликом исходного шума (в опорном канале). Поэтому амплитудный спектр шума в основном канале может быть представлен следующим образом:

$$|W(f, k)| = \sum_{m=0}^M H_m(f) |R(f, k - m)|$$

Коэффициенты передаточных функций $\hat{H}_m(f)$ оцениваются рекурсивно (от кадра к кадру) на основе следующего алгоритма [9]:

$$\hat{H}_m(f, k + 1) = \hat{H}_m(f, k) + \mu [|X(f, k)| / |R(f, k)| - \hat{H}_m(f, k)], \quad m = 0, \dots, M$$

$$|\hat{W}(f, k)| = \sum_{m=0}^M \hat{H}_m(f, k) |R(f, k - m)|$$

Тогда оценка амплитудного спектра полезного сигнала вычисляется на основе спектрального вычитания:

$$|\hat{S}(f, k)| = |X(f, k)| - |\hat{W}(f, k)| = G(f, k) |X(f, k)|,$$

где $G(f, k) = (|X(f, k)| - |\hat{W}(f, k)|) / |X(f, k)|$ – передаточная функция фильтра спектрального вычитания.

Оценка комплексного спектра полезного сигнала вычисляется по комплексному спектру зашумленного сигнала

$$\hat{S}(f, k) = G(f, k) X(f, k)$$

Оценка значений сигнала на кадре вычисляется переходом из частотного представления во временную область:

$$\hat{s}(t) = IDFT\{\hat{S}(f, k)\}$$

Задание 1

Создание опорного канала для стереофильтрации.

Откройте файл *Музыкальная помеха.wav*.

Нажмите **Редактирование** → **Создать опорный канал**.

Укажите файл для фильтрации – основной канал (*Музыкальная помеха.wav*). Нажмите кнопку **Далее**.

Выберите файл для опорного канала (*Опорный сигнал.wav*). Нажмите кнопку **Далее**.

Установите метку начала и конца сигнала в основном канале и соответствующие им метки в опорном канале. Нажмите кнопку **Далее**.

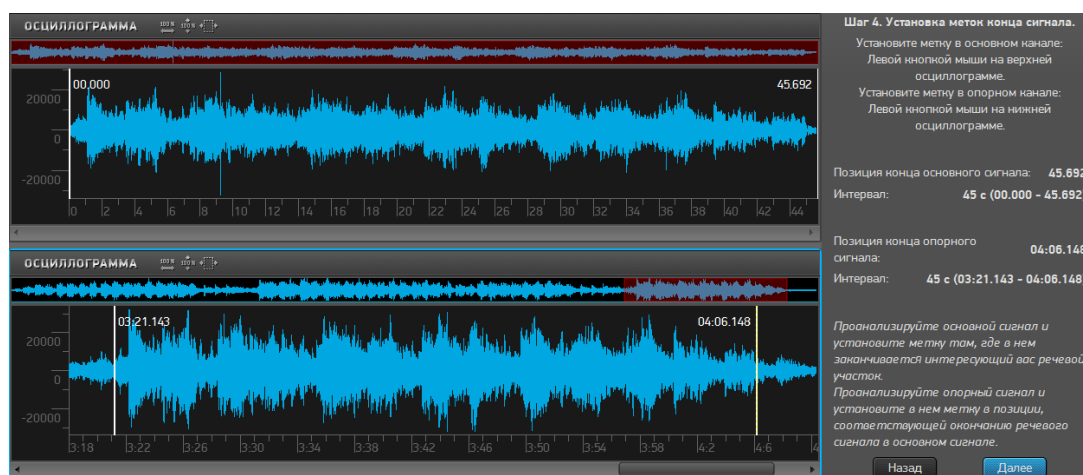


Рисунок 27 – Интерфейс настройки опорного канала

Нажмите кнопку **Завершить**, чтобы создать стереофайл.

Подавление опорного канала

Добавьте фильтр **Подавление опорного канала**.

Выберите «**Левый**» в качестве «Основного канала».

Отрегулируйте «**Глубину подавления**» для достижения наилучшего результата.

Сохраните результат.

Задание 2

Откройте файл *Signal_stereo.wav*.

Осуществите самостоятельную обработку файла стереофильтром.

Сохраните результат.

ЛАБОРАТОРНАЯ РАБОТА № 10

Комплексное применение фильтров

Цель работы: получение навыков использования автоматического фильтра, построения цепочек фильтров.

Задание 1

Откройте любой из файлов: *Track04B.wav*, *overflow.wav*, *airplane.wav*, *columbia.wav*, *buzz.wav*, *Restaurant.wav*, *60Hz.wav*.

Осуществите самостоятельную обработку файла «Автоматическим фильтром».

Сохраните результат.

Осуществите шумоочистку для всех предложенных файлов любым набором фильтров.

Сохраните результаты.

В отчете укажите следующую информацию:

- виды шумов в звуковом файле;
- последовательность выбранных фильтров;
- оценка качества полученного звукового файла по отношению к исходному звуковому файлу;
- предложения по совершенствованию инструментов программы.

СПИСОК ЛИТЕРАТУРЫ

1 Sound Cleaner II. Программное обеспечение для шумоочистки звука STC-S630. Руководство пользователя. ЦБАУ.00564-01 90.

2 Загуменнов А. П. Компьютерная обработка звука. М.: ДМК, 1999, 384 с.

3 Boll S. Suppression of acoustic Noise in Speech Using Spectral Subtraction // IEEE Trans. ASSP, vol.27, N4, 1979, pp.113-120.

4 Bitzer J, Brandt M. Speech Enhancement by Adaptive Noise Cancellation: Problems, Algorithms, and Limits // Trans. of International Conference AES 39, pp.106-113.

5 Уидроу Б., Стирнз С. Адаптивная обработка сигналов. М.: Радио и связь, 1989, 440 с.

6 Lebart K., Boucher J. M. A New Method Based on Spectral Subtraction for Speech Dereverberation // Acta acustica Vol. 87 (2001) pp. 359 – 366.

7 Wang L., Odani K. and Kai A. Dereverberation and denoising based on generalized spectral subtraction by multi-channel LMS algorithm using a small-scale microphone array // EURASIP Journal on Advances in Signal Processing 2012:12.

8 Boll S., and Pulsipher D. Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation // IEEE Trans. ASSP-28, 1980, pp.752-753.

9 Nasu Y., Shinoda K., Furui S. Cross-channel spectral subtraction for meeting speech recognition // Proc. ICASSP, 2011. pp. 4812—4815.

ПРИЛОЖЕНИЕ 1
Пример оформления титульного листа

**САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**

Кафедра речевых информационных систем

ЦИФРОВАЯ ОБРАБОТКА РЕЧЕВЫХ СИГНАЛОВ

ЛАБОРАТОРНАЯ РАБОТА № 1

«Основы работы с программой «SIS»»

Выполнил:
Магистр группы № nnnn: _____

Проверил:
к.т.н., доцент Столбов М. Б.

Санкт-Петербург
2016

ПРИЛОЖЕНИЕ 2

Рекомендации по выполнению лабораторных работ

Скопируйте на свой компьютер каталоги с заданиями (лабораторными работами).

Результаты выполнения работ сохраняйте на своем компьютере или в разделе \\zstorage.stc\RIS\Курсы лекций\Столбов Цифровая обработка речевых сигналов\Лабораторные работы SIS II\ в каталогах со своими фамилиями.

Изучите краткие теоретические сведения и пользуйтесь ими при выполнении заданий.

Результаты выполнения заданий должны быть представлены на рисунках (с подписями внизу под рисунками). Указывайте значения параметров (анализа, обработки сигналов), с которыми вы получили ваши результаты.

Результаты выполнения заданий должны быть прокомментированы (даны ваши научные выводы).

Подписи под рисунками и комментарии необходимо делать в форме, доступной для понимания другим человеком (читателем).

При сдаче работ преподавателю вы должны уметь воспроизвести описанные вами эксперименты и полученные результаты.

Миссия университета – генерация передовых знаний, внедрение инновационных разработок и подготовка элитных кадров, способных действовать в условиях быстро меняющегося мира и обеспечивать опережающее развитие науки, технологий и других областей для содействия решению актуальных задач.

КАФЕДРА РЕЧЕВЫХ ИНФОРМАЦИОННЫХ СИСТЕМ

Кафедра речевых информационных систем (РИС) была создана в 2011 году на факультете Информационных технологий и программирования (ФИТиП). Организатором создания кафедры выступил «Центр речевых технологий». Заведующий кафедрой – доктор технических наук Матвеев Юрий Николаевич.

В рамках приоритетного направления 09.04.02 «Информационные системы и технологии» кафедра РИС осуществляет прием в магистратуру по следующим образовательным программам:

- практико-ориентированная программа «Системы голосового самообслуживания» (с 2015 года);
- образовательная программа «Речевые информационные системы» (с 2011 года);
- совместная образовательная программа «Речевые информационные системы» с Костанайским государственным университетом имени А. Байтурсынова, республика Казахстан с (с 2014 года).

Столбов Михаил Борисович, Кассу Ахмад-Рамез

Цифровая обработка речевых сигналов

Учебно-методическое пособие по лабораторному практикуму

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

**Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49**