

СОДЕРЖАНИЕ

Введение	7
1 Обзор существующих методов повышения разборчивости речи на основе модели слухового восприятия	10
1.1 Повышение разборчивости речи в шумных реверберационных средах, основанный на увеличении энергии речевого сигнала и «обратной» фильтрации	10
1.2 Метод повышения разборчивости речи на основе новой методики уменьшения частоты для пользователей слуховых аппаратов, говорящих на северокитайском языке	12
1.3 Метод повышения разборчивости речи на основе удаления шума с помощью метода адаптивной фильтрации для людей страдающих нейросенсорной тугоухостью	14
1.4 Моделирование потери слуха с использованием компрессионных слуховых гаммачирп-фильтров	15
2 Анализ технического задания.....	17
2.1 Анализ требований к алгоритму.....	17
2.2 Анализ требований к программной реализации	17
2.3 Выбор и обоснование метода решения задачи	17
3 Разработка структуры и алгоритма работы системы повышения разборчивости речи.	19
3.1 Слух и слуховое восприятие	19
3.1.1 Строение слуховой системы	19
3.1.2 Возникновение слуховых ощущений	22
3.1.3 Влияние акустических условий на восприятие речи	23
3.2 Аудиометрия и аудиограмма	24
3.3 Принцип работы системы повышения разборчивости речи	26
3.3.1 Банк гамматон-фильтров.....	26
3.3.2 Настройка компрессора динамического диапазона (КДД)	30
3.3.3 Реализация КДД.....	33
3.4 Принцип работы системы моделирования потери слуха	35
4 Программная реализация системы повышения разборчивости речи	37
4.1 Описание пакета MATLAB.....	37
4.2 Программная реализация системы повышения разборчивости речи	38
4.3 Описание вспомогательных функций	39
4.3.1 Функция gammatoneFast.....	39
4.3.2 Функция bark2frq	40
4.3.3 Функция ttest	40
4.3.4 Функция iso226_rev	42

5 Технико-экономическое обоснование целесообразности проектирования и внедрения системы повышения разборчивости речи.....	43
5.1 Характеристика системы повышения разборчивости речи	43
5.2 Составление плана и определение трудоемкости выполнения научно-исследовательской работы.....	43
5.3 Расчёт цены научно-технической продукции	46
5.3.1 Расчёт затрат по статье «Топливо-энергетические ресурсы для научно-экспериментальных целей»	46
5.3.2 Расчёт затрат по статье «Спецоборудование для научных (экспериментальных) работ»	47
5.3.3 Расчёт затрат по статье «Основная заработная плата научно-технического персонала»	48
5.3.4 Дополнительная заработная плата научно-технического персонала	48
5.3.5 Отчисления на социальные нужды (отчисления в фонд социальной защиты населения и обязательное страхование).....	48
5.3.6 Накладные расходы	49
5.3.7 Полная себестоимость	49
5.4 Оценка уровня качества научно-технического результата	49
6 Анализ результатов тестирования системы.....	52
6.1 Подготовка данных для проведения эксперимента	52
6.2 Описание эксперимента	54
6.3 Результаты эксперимента.....	55
Заключение	57
Список использованных источников	59
Приложение А	61
А.1 Программная реализация алгоритм системы повышения разборчивости речи	61
А.1 Программная реализация алгоритм определения значимости результатов	67
Приложение Б.....	69

фильтров при нормальном слухе, при второй степени потери слуха и при третьей степени потери слуха аппроксимировать значениями в дБ для каждой субполосы. Аппроксимированные значения порогов комфорта представлены в таблице 3.2

Таблица 3.2– Значения для порогов комфорта

№	Порог комфорта при нормальном слухе	Порог комфорта при второй степени потери слуха	Порог комфорта при третьей степени потери слуха
1.	-40.2	-17.6	-7.6
2.	-36.1	-16.3	-5.9
3.	-39.1	-22.1	-10.5
4.	-37.5	-19.5	-5.2
5.	-36.5	-16.9	-7.8
6.	-37.2	-21.2	-11.9
7.	-42.4	-17.4	-9.3
8.	-39	-22.8	-5.5
9.	-40.6	-14.4	-11.7
10.	-41.8	-18.1	-11.7
11.	-41.4	-16.5	-10.9
12.	-37	-18.9	-9.7
13.	-34.4	-15.1	-4.9
14.	-42.6	-18.5	-9.6
15.	-41.4	-22.5	-11.4
16.	-41.4	-15.6	-8
17.	-41.6	-15.9	-7.9
18.	-40.5	-24.7	-11.8
19.	-41.5	-22.1	-11
20.	-39.3	-17.8	-9.7

Блок-схема алгоритма настройки компрессора динамического диапазона сигнала представлена на рисунке 3.7.

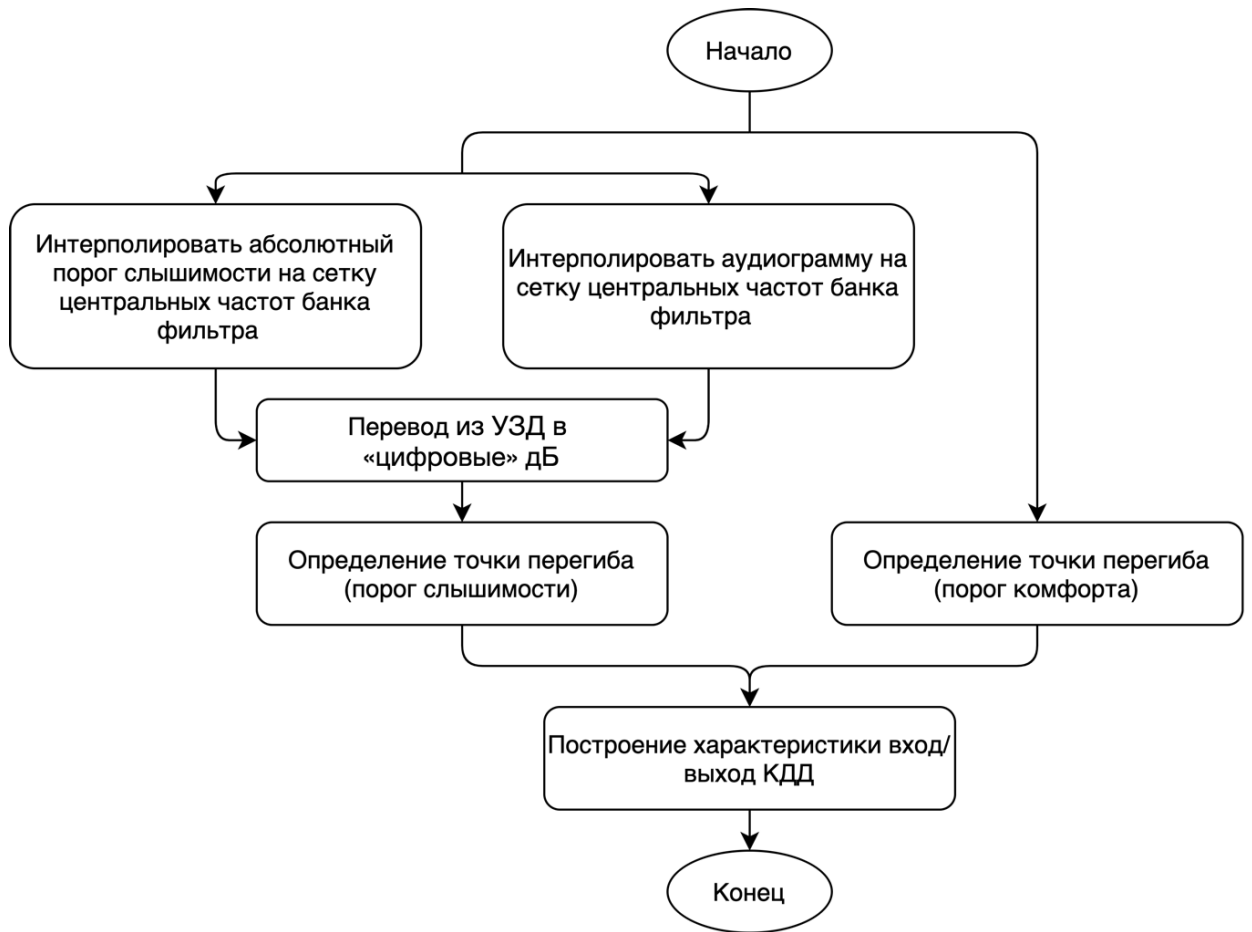


Рисунок 3.7 – Блок-схема алгоритма настройки КДД

Настройка системы повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала по аудиограмме представлена на плакате ГУИР 460818.004 ПЛ к данной работе. На плакате отображены абсолютный порог слышимости и порог слышимости в норме, характеристика КДД для центральной частоты равной 1172 Гц, аудиограмма человека со второй степенью потери слуха, амплитудно-частотная характеристика для банка гамматон-фильтров и их взаимосвязь для определения точки порога слышимости для настройки системы.

3.3.3 Реализация КДД

Для реализации КДД в каждой субполосе на выходе банка фильтров, на частотах f_C производится оценка мощности сигнала:

$$P_k^{in}[n] = \begin{cases} P_k[n-1]\alpha + x_k^2[n](1-\alpha), & \text{если } x_k^2[n] > P_k[n-1] \\ P_k[n-1]\beta + x_k^2[n](1-\beta), & \text{если } x_k^2[n] \leq P_k[n-1] \end{cases}, \quad (3.1)$$

$$k = 1, 2 \dots K,$$

Алгоритм такой проверки заключается в том, что для каждого участника эксперимента вычисляется разница между разборчивостью речи с применением и без применения метода коррекции слуха. При этом нулевая гипотеза подразумевает то, что истинная разность равна нулю. Для проверки этой гипотезы в среде MATLAB использовался t-критерий Стьюдента для вычисления р-значения с установленным пороговым уровнем значимости α который в данной работе был выбран $\alpha = 0.05$. Нулевая гипотеза отвергается, если р-значение меньше уровня значимости α [22].

Описание параметров данной функции представлено в таблице 4.3., а описание возвращаемых значений представлено в таблице 4.4.

Таблица 4.3– Описание параметров функции *ttest*

Параметры	Возможные значения	Описание
x	Массив чисел	Первая выборка значений
y	Массив чисел	Вторая выборка значений
<i>name</i>	Строка	Название параметра, который задается следующим параметром <i>value</i>
<i>value</i>	Число	Пороговый уровень значимости

Таблица 4.4– Описание возвращаемых значений функции *ttest*

Возвращаемые значения	Возможные значения	Описание
h	Число	Результат для нулевой гипотезы
p	Число	Уровень значимости
ci	Массив из двух чисел	Доверительный интервал
<i>stats</i>	Структура со следующими полями: <i>tstat</i> (число), <i>df</i> : (число), <i>sd</i> (число)	Информацию о статистике теста

Исходный код программной реализации проверки приведен в приложении А.2.

5 ТЕХНИКО-ЭКОНОМИЧЕСКОЕ ОБОСНОВАНИЕ ЦЕЛЕСООБРАЗНОСТИ ПРОЕКТИРОВАНИЯ И ВНЕДРЕНИЯ СИСТЕМЫ ПОВЫШЕНИЯ РАЗБОРЧИВОСТИ РЕЧИ

5.1 Характеристика системы повышения разборчивости речи

Проводимая научно-исследовательская работа заключается в проектировании системы повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала.

По состоянию на 2021 год более 430 миллионов человек нуждаются в реабилитации для решения проблемы потери слуха. По прогнозам, к 2050 г. по меньшей мере 700 миллионов человек будут нуждаться в реабилитационных услугах в связи с потерей слуха и около 2,5 миллиарда человек в той или иной степени будут страдать от проблем со слухом. В связи с этим существует задача оперативной коррекции слуховых патологий.

Главная область применения разработанной системы – слуховые аппараты. Основная цель системы повышения разборчивости речи заключается в непосредственном усилении входного сигнала и повышения разборчивости в соответствии с характеристикой слуха тугоухого человека. Важной особенностью системы является применение в ней банка гамматон-фильтров, имитирующего механизм частотной декомпозиции звука в ухе. Другой особенностью системы является её автоматическая настройка по аудиограмме слабослышащего человека. Эти особенности помогают разработать систему, которая справляется с задачей повышения разборчивости лучше, чем ряд существующих методов.

5.2 Составление плана и определение трудоемкости выполнения научно-исследовательской работы

Данная работа проводится в рамках внутреннего исследования университета с дальнейшим развитием в научных статьях и журналах. План на проведение научно-исследовательской работы представлен в таблице 5.1. В данном плане отражены этапы проектирования системы повышения разборчивости речи на основе модели слухового восприятия и компрессии динамического диапазона сигнала, а также количество и состав исполнителей: научный руководитель и ответственный исполнитель.

Основными методами определения трудоёмкости являются система аналогов, метод прямого счёта и метод экспертных оценок. В данной работе при расчете трудоемкости использовался метод прямого счета. При прямом

низких частотах использованный банк фильтров имеет длинную импульсную характеристику, что увеличивает количество операций с сигналом.

Настройка КДД выполнена согласно аудиограммам для второй и третьей степени потери слуха.

Для проверки работоспособности системы выполнено её MATLAB-моделирование. В качестве входного сигнала использовался речевой сигнал, спектрограммы и временное представление которого представлены в графическом документе ГУИР 460818.005 ПЛ к данной работе. Кроме того, на плакате отражены спектрограммы сигнала, обработанного методом модерирования потери слуха и методом повышения разборчивости речи для второй степени потери слуха. Этот сигнал является пятнадцатисекундным отрезком аудиодорожки (таблица 1), которую давали прослушать участникам эксперимента. Данные спектрограммы объясняют результаты эксперимента, которые также представлены на данном плакате в виде гистограмм.

Из результатов, отраженных на плакате видно, что используемая система КДД выполняет частотно-зависимое усиление, что, в свою очередь, приводит к увеличению уровня исходного звукового сигнала и положительно влияет на разборчивость речи. Искажения, появляющиеся в результате обработки, не влияют на общий уровень разборчивости речи и возможность идентифицировать говорящего человека.

Стоит отметить, что существует возможность проведения усовершенствования полученной системы в результате проектирования, например путем добавления дополнительной точки перегиба (болевого порога) для характеристики вход/выход КДД. Это позволит избежать такой формы искажения звука, как клиппинг.

ПРИЛОЖЕНИЕ А

(обязательное)

Код программы

А.1 Программная реализация алгоритм системы повышения разборчивости речи

```
close all; clc;

% Инициализация констант
K = 21;
frame_size = 256;
coefficient_a = 0.9900; % attack time coefficient
coefficient_b = 0.9992; % release time coefficient
freq_aud = [125 250 500 750 1000 2000 3000 4000 6000 8000]; % сетка частот
% HT = [24 30 36 35 34 40 42 44 41 40]; % 1я степень тугоухости
HT = [43 50 50 47 45 50 52 53 58 60]; % 2я степень тугоухости
% HT = [53 66 60 62 64 54 60 67 69 70]; % 3я степень тугоухости
% HT = [70 75 76 78 80 77 79 80 82 84]; % 4я степень тугоухости
input_signal = zeros(1,2^12);
input_signal(1,1) = 1;
bark_sc=1.5:1:K-0.5;
CL_in_2 = [-40.2,-36.1,-39.1,-37.5,-36.7,-37.2,-42.4,-39,-40.6,-41.8,-41.4,-37,-34.4,-42.6,-41.4,-41.4,-41.6,-40.5,-41.5,-39.3];

CL_out_2 = [-17.6,-16.3,-22.1,-19.5,-16.9,-21.2,-17.4,-22.8,-14.4,-18.1,-16.5,-18.9,-15.1,-18.5,-22.5,-15.6,-15.9,-24.7,-22.1,-17.8];

CL_in_3 = [-40.2,-36.1,-39.1,-37.5,-36.7,-37.2,-42.4,-39,-40.6,-41.8,-41.4,-37,-34.4,-42.6,-41.4,-41.4,-41.6,-40.5,-41.5,-39.3];

CL_out_3 = [-7.6,-5.9,-10.5,-5.2,-7.8,-11.9,-9.3,-5.5,-11.7,-11.7,-10.9,-9.7,-4.9,-9.6,-11.4,-8,-7.9,-11.8,-11,-9.7];

CL_in = CL_in_2;
CL_out = CL_out_2;

% Выбираются центральные частоты банка фильтров в соответствии с
% психоакустической шкалой Барков
fc = bark2frq(bark_sc);
[spl, freq_iso] = iso226_rev(0);

% Считывается входного звукового сигнала
addpath('input_signals');
filename = 'черный ворон 25с';
cond = '_corr_HL3';
[audio, fs] = audioread([filename '.mp3']);

audio_input_signal = audio(:, 1)';

output_signal = gammatoneFast(input_signal, fc);
create_gammaton_bank_filter_figure(output_signal, K, fs, freq_aud);

from_digdb_to_spl = 105;

% Нахождение порога слышимости
interped_HT = interp1(freq_aud, HT, fc, 'linear');
interped_spl = interp1(freq_iso, spl-from_digdb_to_spl, fc, 'linear');
hi = interped_HT+interped_spl;
```



```

% Построение графиков аудиограммы, порогов слышимости в норме и при наличии
% патологии
create_hearing_thresholds_figure(freq_aud, HT, hi, freq_iso, spl, fc);

% Происходит разложение входного сигнала на субполосы
audio_output_signal = gammatoneFast(audio_input_signal, fc);

% Производится коррекция амплитуд
corrected_audio_output_signal = audio_output_signal.*1.8;

% Инициализация переменных
N_frames=floor(length(corrected_audio_output_signal(1,:))/frame_size);
p_start = repmat(2^(-10),length(1:K-1),1);
g_begin = ones(length(1:K-1));
synthesized_audio_output_signal = zeros(1, N_frames * frame_size);
processed_audio_output_signal = zeros(K, N_frames * frame_size);
y = zeros(1, N_frames * frame_size);

% Конфигурация графика компрессора динамического диапазона
figure;
configure_figure_settings('Компрессор динамического диапазона', 'Входная
мощность, дБ', 'Выходная мощность, дБ');
hold on;
grid on;

for M=1:K-1
    % Определение точек перегиба компрессора динамического диапазона
    drc_in = [-120 interped_spl(1,M) CL_in(M) 0];
    drc_out = [-120 hi(1,M) CL_out(M) 0];
    % Построение графика компрессора динамического диапазона
    plot (drc_in, drc_out, 'k', 'LineWidth', 1);
    plot (drc_in, drc_out, 'ok', 'LineWidth', 1.5);

    % Реализация алгоритма нахождения коэффициентов усиления для каждой
    % субполосы
    for N=1:N_frames
        i_beg = (N-1)*frame_size+1;
        i_end = N*frame_size;
        x_frame = corrected_audio_output_signal(M, i_beg : i_end);

        % Разбиение каждой субполосы на фреймы для оптимизации алгоритма
        for j=1:frame_size
            % Производится оценка уровня мощности
            p_cur = x_frame(j)^2;

            if p_cur > p_start(M)
                p_start(M) = p_start (M) * coefficient_a + p_cur * (1 -
coefficient_a);
            else
                p_start(M) = p_start (M) * coefficient_b + p_cur * (1 -
coefficient_b);
            end
        end

        p_end = p_start (M);
        % Полученный уровень мощности переводится в дБ
        p_end_db = 10*log10(p_end);
        % Вычисляется коэффициент усиления в каждой субполосе
        p_out_db = interp1(drc_in, drc_out, [p_end_db], 'linear');
        g_end_db = p_out_db - p_end_db;
        % Коэффициент усиления переводится из дБ в линейный масштаб
        g_end = 10^(g_end_db / 20);
        g_line = interp1([0 frame_size], [g_begin(M) g_end],1:frame_size,
'linear');
    end
end

```

```

        g_begin(M) = g_end;
        % Производится усиления
        y(i_beg:i_end) = x_frame.*g_line;
    end

    % Синтез выходного сигнала
    synthesized_audio_output_signal = synthesized_audio_output_signal + y;
    processed_audio_output_signal(M, :) = y;
end

create_characteristic_drc_input_output_figure(drc_in, drc_out);
create_input_output_subband_signals_figure(corrected_audio_output_signal,
processed_audio_output_signal);

% Построение графиков временного представления и спектрограмм входного и
выходного сигнала
create_input_output_signal_figure(audio_input_signal,
synthesized_audio_output_signal, fs, freq_aud);

% Запись выходного звукового сигнала
audiowrite([filename cond '.wav'], synthesized_audio_output_signal, fs);

function create_gammaton_bank_filter_figure(corrected_output_signal, K, fs,
freq_aud)
    figure;
    hold on;
    set(gca, 'YLim', [-90 0]);

    configure_figure_settings('', 'Частота, Гц', 'Амплитуда, дБ');

    for N=1:K-1
        [h, w] = freqz(corrected_output_signal(N,:), 1, 4096);
        hDb = 20*log10(h);
        wHz = w/pi*(fs/2);
        plot(wHz, hDb, 'LineWidth', 2.5);
    end
    grid on;
    ylim ([-72 0]);
end

function create_hearing_thresholds_figure(freq_aud, HT, hi, freq_iso, spl,
fc)
    figure;
    array = 1:length(freq_aud);
    hold on;
    plot (array, -1*HT, 'k', 'LineWidth', 2);
    ylim ([-70 10]);
    xlim ([array(1) array(end)]);
    plot (array, -1*HT, 'ok', 'LineWidth', 2);
    grid on;
    set(gca, 'YTickLabel', [70:-10:-10]); %90,80,70,60,50,40,30,20,10,0,-10
    set(gca, 'XTickLabel', freq_aud);
    set(gca, 'XTick', array(1):array(end));
    set(gca, 'Clim', [-65 15]);
    configure_figure_settings('Аудиограмма при 2-ой степени тугоухости',
'Частота, Гц', 'Потеря слуха, дБ');

    figure;
    plot(freq_iso, spl-80, 'k', 'LineWidth', 2);
    configure_figure_settings('Абсолютный порог слышимости', 'Частота, Гц',
'Амплитуда, дБ');
    xlim ([freq_iso(1) freq_iso(end)]);
    ylim ([-90 0]);
    grid on;

```

```

figure;
plot(fc, hi, 'k', 'LineWidth', 2);
configure_figure_settings('Порог слышимости при патологии', 'Частота,
Гц', 'Амплитуда, дБ');
xlim ([fc(1) fc(end)]);
ylim ([-90 0]);
grid on;
end

function configure_figure_settings(titleSpecgram, xlabelSpecgram,
ylabelSpecgram)
set(gca, 'Clim', [-65 15]);
title(titleSpecgram);
xlabel(xlabelSpecgram);
ylabel(ylabelSpecgram);
set(gca, 'FontName', 'Times New Roman');
set(gca, 'FontSize', 18);
end

function create_characteristic_drc_input_output_figure(drc_in, drc_out)
% Передаточная характеристика устройства
figure;
plot(drc_in, drc_out);
grid on;
configure_figure_settings('DRC Characteristic', 'Input magnitude, dB',
'Output magnitude, dB');
xlim([-90 0]);
ylim([-90 0]);
end

function create_input_output_signal_figure(audio_input_signal,
synthesized_audio_output_signal, fs, freq_aud)
figure;

subplot(2, 2, 1);
specgram(audio_input_signal, 512, fs, kaiser(512,7),475);
ylim ([0 freq_aud(end)]);
configure_figure_settings('Спектрограмма входного сигнала', 'Время, с',
'Частота, Гц');

subplot(2, 2, 2);
specgram(synthesized_audio_output_signal, 512, fs, kaiser(512,7),475);
ylim ([0 freq_aud(end)]);
configure_figure_settings('Спектрограмма выходного сигнала', 'Время, с',
'Частота, Гц');

subplot(2, 2, 3);
plot((1:length(audio_input_signal)) / fs, audio_input_signal);
ylim ([-2 2]);
xlim ([0 25]);
configure_figure_settings('Временное представление сигнала', 'Время, с',
'Амплитуда, дБ');

subplot(2, 2, 4);
plot((1:length(synthesized_audio_output_signal)) / fs,
synthesized_audio_output_signal);
ylim ([-2 2]);
xlim ([0 25]);
configure_figure_settings('Временное представление сигнала', 'Время, с',
'Амплитуда, дБ');
end

```

```

function
create_input_output_subband_signals_figure(corrected_audio_output_signal,
processed_audio_output_signal)
    figure;
    subplot(2,1,1);
    plot(corrected_audio_output_signal);
    configure_figure_settings('Input signal', 'Time, s', 'Magnitude, dB');
    ylim([-1 1]);

    subplot(2,1,2);
    plot(processed_audio_output_signal);
    configure_figure_settings('Output signal', 'Time, s', 'Magnitude, dB');
    ylim([-1 1]);
end

function create_gammaton_bank_filter_figure(corrected_output_signal, K, fs,
freq_aud)
    figure;
    hold on;
    set(gca, 'YLim', [-90 0]);

    configure_figure_settings('', 'Частота, Гц', 'Амплитуда, дБ');

    for N=1:K-1
        [h, w] = freqz(corrected_output_signal(N,:),1,4096);
        hDb = 20*log10(h);
        wHz = w/pi*(fs/2);
        plot(wHz, hDb, 'LineWidth',2.5);
    end
    grid on;
    ylim ([-72 0]);
end

function f = bark2frq(b)
f=52547.6*(26.28-b-(22.11-1.1*b).*(b>20.1)/6.1-(3*b-6).*(b<2)/17).^(-1)-1960;

function [bm,env,delay] = gammatoneFast(x,cfs,fs,align)
    if nargin < 3
        fs = 16000; % default sampling frequency
    end
    if nargin < 4
        align = false; % default phase alignment
    end

    % check inputs
    assert(isvector(x) & isnumeric(x), 'iosr:gammatoneFast:invalidX', 'x must
be a vector')
    assert(isvector(cfs) & isnumeric(cfs), 'iosr:gammatoneFast:invalidCfs',
'cfs must be a vector')
    assert(isscalar(fs), 'iosr:gammatoneFast:invalidFs', 'fs must be a
scalar')
    assert(islogical(align) & numel(align)==1,
'iosr:gammatoneFast:invalidAlign', 'align must be logical')

    % number of frequency channels
    numchans = length(cfs);

    filterOrder = 4; % filter order
    gL = 2^nextpow2(0.128*fs); % gammatone filter length at least 128 ms
    b = 1.019.*24.7.*(4.37.*cfs./1000+1); % rate of decay or bandwidth

    gt = zeros(gL,numchans); % Initialise IR
    tc = zeros(size(cfs)); % Initialise time lead
    phase = 0;

```

```

tpt=(2*pi)/fs;
gain=((1.019.*b.*tpt).^filterOrder)./6; % based on integral of impulse

tmp_t = (0:gL-1)/fs;

% calculate impulse response
for i = 1:numchans
    if align
        tc(i) = (filterOrder-1)./(2*pi*b(i));
        phase = -2*pi*cfs(i)*tc(i);
    end
    gt(:,i) = gain(i)*fs^3*tmp_t.^(filterOrder-1).*exp(-
2*pi*b(i)*tmp_t).*cos(2*pi*cfs(i)*tmp_t+phase);
end

% if input is row vector, transpose to column vector
rot = false;
if size(x,1)==1
    x = x';
    rot = true;
end

% gammatone filtering using FFTFILT
bm = fftfilt(gt, repmat(x,1,numchans));

% Hilbert envelope
env = abs(hilbert(bm));

% delay due to time lead
delay = round(tc.*fs);

% remove time lead
for i = 1:numchans
    bm(:,i) = [bm(delay(i)+1:end,i); zeros(delay(i),1)];
    env(:,i) = [env(delay(i)+1:end,i); zeros(delay(i),1)];
end

% transpose output if necessary
if rot
    bm = bm';
    env = env';
end

end

```