УДК 616.71

ИССЛЕДОВАНИЕ МЕТОДОВ СЕГМЕНТАЦИИ ГОЛОСОВЫХ СИГНАЛОВ

Пастернак В. В., Стецко В.Ю., студенты гр.950702

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Вашкевич М.И. – канд. техн. наук

Аннотация. В работе рассмотрены методы сегментации голосовых сигналов, а именно Waveform Matching метод и Peak-picking метод. Результатом работы этих методов является извлечение частоты основного тона голосового сигнала. Методы проверялись на тестовых сигналах, имеющих синусоидальную форму и различную частотную модуляцию. Для проверки производительности методы сравнивались по отношению параметров частотной пертурбации к их теоретическим значениям.

Ключевые слова. Сегментация голосовых сигналов, частота основного тона.

Введение

Меры искажения голоса, такие как джиттер и шиммер, зависят от точного извлечения частоты основного тона из аудиосигнала. Метод сегментации, а точнее, его способность точно определить цикличные границы напрямую влияет на точность этих измерений, особенно если в сигнале присутствуют шум и модуляция. В связи с этим весьма актуальна задача повышения точности работы методов сегментации.

В данной работе исследуются два метода сегментации голосовых сигналов [1]:
– метод отбора локальных максимумов (англ. *PP – peak-peaking method*);
– метод подгонки формы сигнала (англ. *WM – waveform matching method*).

Грубая размета сигнала на периоды основного тона

Входной информацией для обоих методов является грубая разметка сигнала на периоды основного тона. Грубая разметка задается последовательностью $I_N(i)$, $i = 1,2, \dots N_c + 1$, где N_c – число периодов основного тона в исходном сигнале. Значение $I_N(i)$ определяет границу между $(i-1)$-м и *i*-м периодом основного тона. Для получения разметки $I_N(i)$ анализируемый сигнал $x(n)$ пропускается через КИХ-фильтр нижних частот с частотой среза:

$$F_c = 1{,}5 \cdot F_0 \quad (1)$$

где F_0 – грубая оценка частоты основного тона.

Фильтрация необходима, чтобы удалить из сигнала все гармонические компоненты за исключением основной гармоники. Для определения F_0 из входного сигнала выбирался фрагмент голосового сигнала длительностью 50 мс (рисунок 1), от которого затем вычислялась автокорреляционная функция (АКФ) (рисунок 2). Положение первого пика АКФ для $\tau > 0$ определяет задержку соответствующую периоду основного тона T_0. Далее частота основного тона определялась как $F_0 = 1/T_0$.

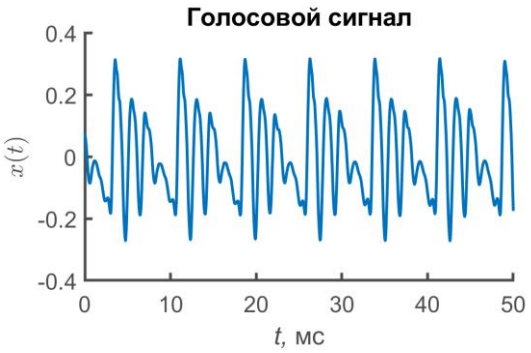

Рисунок 1 – Пример голосового сигнала

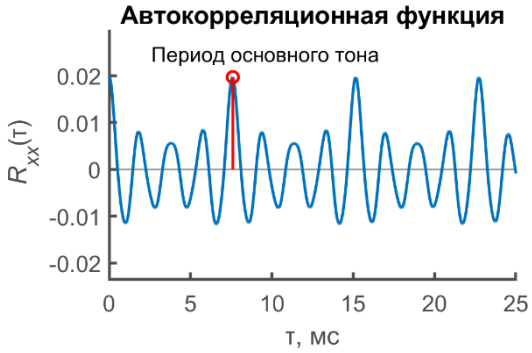

Рисунок 2 – Автокорреляционная функция голосового сигнала

После фильтрации определяются точки пересечения нуля из положительного значения сигнала в отрицательное. Убрав задержку, вызванную фильтром, эти точки можно перенести на исходный сигнал и использовать их в качестве грубой разметки $I_N(i)$. На рисунке 3 показан фильтрованный сигнал с полученной грубой разметкой $I_N(i)$. А на рисунке 4 показан исходный сигнал с нанесенной грубой разметкой. Можно видеть, что полученная указанным образом грубая разметка позволяет эффективно определить приблизительные границы периодов основного тона.

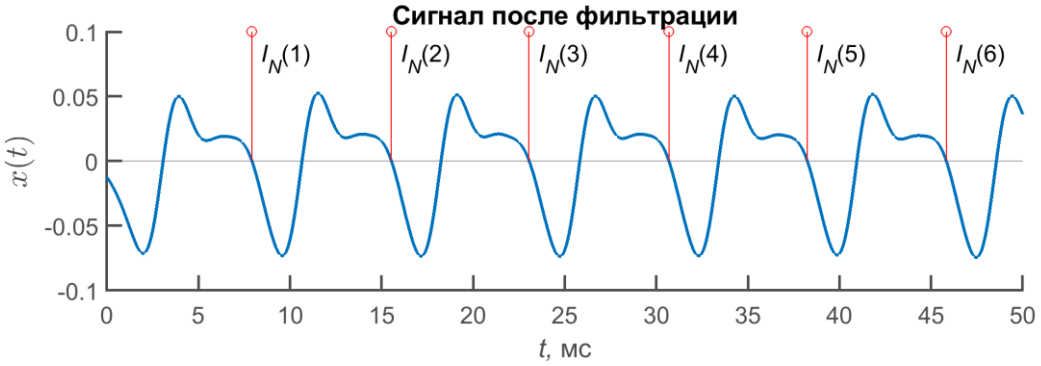

Рисунок 3 – Фильтрованный с грубой разметкой

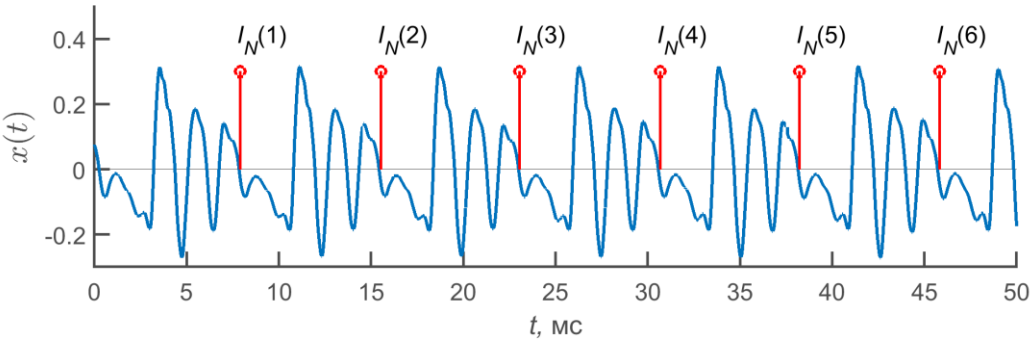

Рисунок 4 –Исходный сигнал с грубой разметкой

Метод сегментации путем подгонки формы сигнала

Суть метода *подгонки формы сигнала* состоит в поиске минимумов между точками грубой разметки так, чтобы среднеквадратичная ошибка между двумя соседними формами сигнала была минимальной, чтобы далее найти частоту основного тона, используя формулу:

$$F(i-1) = \frac{f_s}{P(i) - P(i-1) + \delta}, \quad (2)$$

где f_s – частота дискретизации сигнала.

Полная процедура метода заключается в следующем:

1. Находим местоположение абсолютного минимума – точку $P(1)$, которая находится между первыми двумя точками грубой разметки – $I_N(1)$ и $I_N(2)$.
2. Далее создаем цикл для $i = 2$ до $N_C + 1$ и проделываем операции внутри цикла.
   А) Находим первоначальное предположение:

$$P(i) = I_N(i) + (P(i-1) - I_N(i-1)) \quad (3)$$

Пример найденных точек $P(i)$ согласно выражению (3) показан на рисунке 5.

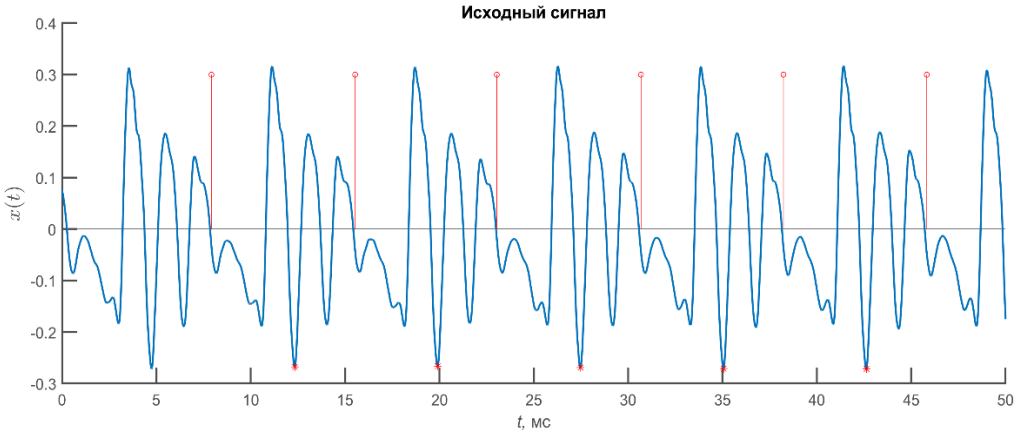

Рисунок 5 – Сигнал с минимумами между точками грубой разметки

Б) Устанавливаем верхние и нижние пределы видимости:

$$J_1 = P(i) - PERC \cdot (P(i) - P(i-1)) \quad (4)$$

$$J_2 = P(i) + PERC \cdot (P(i) - P(i-1)) \quad (5)$$

где $PERC$ – заданный параметр. Например, 0.05 – 0.15 (т.е. допускается изменение на 5-15%).

В) Находим точку J_m между пределами видимости J_1 и J_2, чтобы среднеквадратичная ошибка $ERR(J_m)$ была минимальной, где

$$ERR(j) = \frac{1}{j - P(i-1)} \cdot \sum_{k=P(i-1)}^{j-1} (x\left(k + \left(j - P(i-1)\right)\right) - x(k))^2 \quad (6)$$

где $x(n)$ – анализируемый сигнал.

Г) Если $J_m = J_1$ или J_2, расширяем поиск за пределы J_1 или J_2 до тех пор, пока не будет найден минимум функции $ERR(j)$.

Д) Пусть $P(i) = J_m$. Следует обратить внимание, что $P(i)$ – целое число, следовательно, разница $P(i) - P(i-1)$ дает оценку периода основного тона с точностью, ограниченной частотой дискретизации.

Е) Чтобы улучшить оценку частоты, найдем полином второй степени, проходящий через точки $ERR(J_m - 1)$, $ERR(J_m)$ и $ERR(J_m + 1)$, чтобы найти положение точки минимума. Расстояние между точкой минимума и J_m равно:

$$\delta = -0.5 \cdot \frac{ERR(J_m + 1) - ERR(J_m - 1)}{ERR(J_m + 1) - 2 \cdot ERR(J_m) + ERR(J_m - 1)} \quad (7)$$

Ж) Частота *(i-1)*-ого цикла равна:

$$F(i-1) = \frac{f_s}{P(i) - P(i-1) + \delta} \quad (8)$$

Метод подгонки формы сигнала реализован в *Matlab* следующим образом:

```
function [F,F_time] = WM_method(x, fs, I_N)
% x – входной сигнал
% fs – частота дискретизации
% I_N – грубая разметка входного сигнала

PERC = 0.05;
Nc = length(I_N);
F = zeros(1,Nc-3);

[~,P(1)] = min(x(I_N(1):I_N(2)));
P(1) = P(1) + I_N(1);

for i = 2:Nc-2
   P(i) = I_N(i) + (P(i-1) - I_N(i-1));
   J1 = P(i) - round(PERC * (P(i) - P(i-1)));
   J2 = P(i) + round(PERC * (P(i) - P(i-1)));
```

```
function [err_cur] = ERR(j, P, i, buff)

sum = 0;
for k=P(i-1):(j-1)
   sum = sum + (buff(k+(j-P(i-1))) - buff(k))^2;
end
err_cur = (1/(j-P(i-1))) * sum;

end
-----------------------------------------------------------------------------------------
function [err_Jm_plus1, err_Jm_minus1, err_Jm] = ERR_Jm(Jm, P, i, buff)

sum = 0;
for k=P(i-1):(Jm+1-1)
   sum = sum + (buff(k+(Jm+1-P(i-1))) - buff(k))^2;
```

```
  search_complete = false;
  while (~search_complete)
    err_min = inf;
    Jm = [];
    for j = J1:J2
      err_cur = ERR(j, P, i, x);
      if (err_cur<err_min)
        err_min = err_cur;
        Jm = j;
      end
    end
    if (Jm == J1)
      J1 = J1 - round(0.5*(J2-J1));
    elseif (Jm == J2)
      J2 = J2 + round(0.5*(J2-J1));
    else
      search_complete = true;
    end
  end
  P(i) = Jm;
  [err_Jm_plus1, err_Jm_minus1, err_Jm] = ERR_Jm(Jm, P, i, x);
  delta = -0.5 * ((err_Jm_plus1 - err_Jm_minus1)/(err_Jm_plus1 -
2*err_Jm + err_Jm_minus1));
  F(i-1) = fs / (P(i) - P(i-1) + delta);
end
```

```
end
err_Jm_plus1 = (1/(Jm+1-P(i-1))) * sum;

sum = 0;
for k=P(i-1):(Jm-1-1)
  sum = sum + (buff(k+(Jm-1-P(i-1))) - buff(k))^2;
end
err_Jm_minus1 = (1/(Jm-1-P(i-1))) * sum;

sum = 0;
for k=P(i-1):(Jm-1)
  sum = sum + (buff(k+(Jm-P(i-1))) - buff(k))^2;
end
err_Jm = (1/(Jm-P(i-1))) * sum;

end
```

Метод отбора локальных максимумов

Метод *отбора локальных максимумов* использует нахождение отрицательных пиков на каждом фрагменте сигнала между точками грубой разметки. К точке отрицательного пика применяется интерполяция второго порядка. После проделанных действий высчитывается частота основного тона на каждом фрагменте:

$$F(i-1) = \frac{Fs}{P(i) - P(i-1)} \tag{9}$$

где Fs – частота дискретизации, $P(i)$ – местоположение отрицательного пика на каждом фрагменте.

На рисунке 6 показан исходный сигнал с найденными отрицательными пиками между точками грубой разметки.

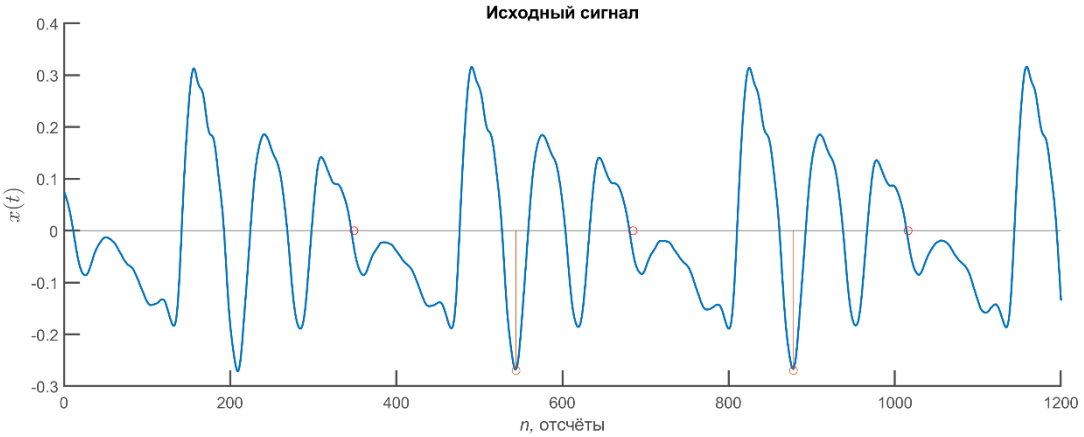

Рисунок 6 – Временное представление сигнала с отрицательными пиками и грубой разметкой

Найденный массив частоты основного тона проверяется по заданному процентному отклонению для нахождения неверной частоты, если она имеется. Данные заносятся в массив логический значений.

Метод отбора локальных максимумов реализован в *Matlab* следующим образом:

```
function [F,ERR,F_time] = PPM(buff, fs, I_N)
% buff – входной сигнал
% fs – частота дискретизации
% I_N – грубая разметка входного сигнала
```

```
PERC = 0.05;
Nc = length(I_N);

for i = 1:Nc - 1
   J1 = I_N(i);
   J2 = I_N(i + 1);
   [~, Pi] = min(buff(J1:J2));
   Pi = Pi + J1 - 1;
   Pr(i) = Pi + (-0.5 * (buff(Pi + 1) - buff(Pi - 1)))/(buff(Pi + 1) - 2 * buff(Pi) +
buff(Pi - 1));

   if (i > 1)
      F(i - 1) =  fs / (Pr(i) - Pr(i - 1));
   end
end

for i = 1:Nc - 3
   ERR(i) = F(i + 1) - F(i) >= F(i) * PERC;
end

end
```

На рисунке 7 представлены результаты поиска частоты основного тона по методам *отбора локальных максимумов* и *подгонки формы сигнала* соответственно.

Рисунок 7 – Частоты основного тона по каждому из методов

Описание эксперимента

Для проведения экспериментов использовались полигармонические сигналы с добавлением амплитудной и частотной модуляции. Для получения полигармонического колебания использовался «смещенный» синус:

$$S_b(\alpha) = \begin{cases} 0{,}5\left(1 - \cos\left(\frac{1+r}{2}\alpha\right)\right), & 0 \le \alpha < \frac{2\pi}{1+r}, \\ 0{,}5\left(1 - \cos\left(\frac{1+r}{2r}\alpha + \frac{r-1}{r}\pi\right)\right), & \frac{2\pi}{1+r} \le \alpha < 2\pi. \end{cases} \quad (10)$$

Степень «смещенности» синуса определяется параметром r, при $r = 1$ получается чистый синус. Входной аргумент α – фазовое значение находящееся в диапазоне от 0 до 2π. Обратите внимание, что если в качестве аргумента в $S_b(\alpha)$ попадет $\alpha > 2\pi$, то необходимо вычесть из него 2π и только затем произвести вычисление по формуле (10). В случае, если после первого вычитания окажется, что α все ещё больше 2π вычитание следует повторить. И так до тех пор, пока α не попадет в диапазон $0 \le \alpha < 2\pi$.

Параметрами частотно-модулированного (ЧМ) сигнала являются: f_0 – частота основного тона (Гц); f_m – частота модуляции (Гц); k_f – индекс модуляции. Для генерирования ЧМ-сигнала использовалось выражение:

$$s(n) = S_b(\alpha(n)), \quad (11)$$

где

$$\alpha(n) = 2\pi \sum_{m=0}^{n} \frac{f_0}{f_s}\left(1 + k_f \cos\left(\frac{2\pi f_m}{f_s} m\right)\right). \quad (12)$$

Параметрами амплитудно-модулированного (АМ) сигнала являются: f_s – частота дискретизации (Гц); f_{am} – частота модуляции (Гц); k_a – индекс модуляции. Для генерирования АМ-сигнала использовалось выражение:

$$s_{AM}(n) = s(n) * \left(1 + k_a \cos\left(\frac{2\pi f_{am}}{f_s} n\right)\right). \quad (13)$$

На рисунке 8 приведен сегмент тестового сигнала длиной 100 миллисекунд после амплитудной модуляции.

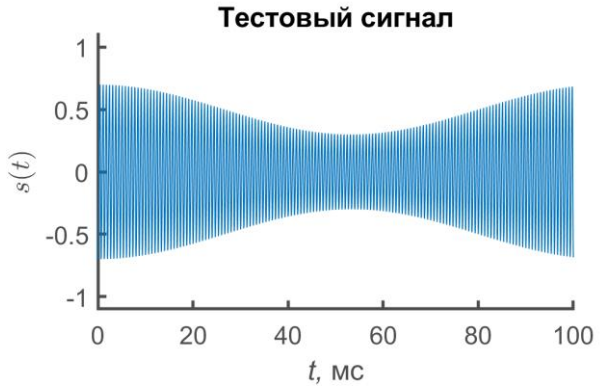

Рисунок 8 – Сегмент тестового сигнала после амплитудной модуляции

Для сравнения производительности между методами сегментации были созданы 20 тестовых сигналов, отличающиеся величиной частотной ($f_0 = 150$ Гц). На рисунке 9 приведен пример контура частоты основного тона (ЧОТ) тестового сигнала, а на рисунке 10 десятимиллисекундный сегмент тестового сигнала.

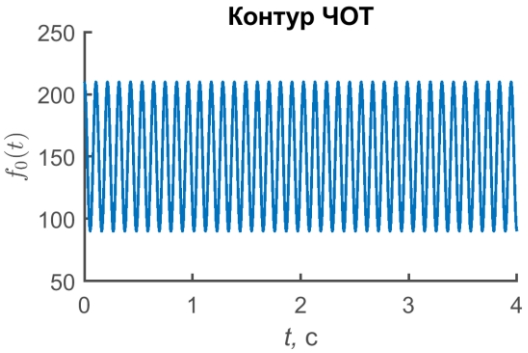

Рисунок 9 – Контура частоты основного тона тестового сигнала

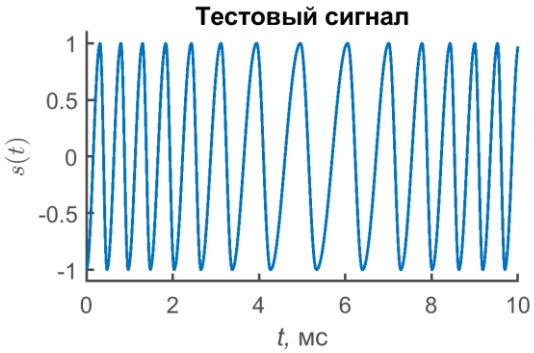

Рисунок 10 – Сегмент тестового сигнала

К каждому тестовому сигналу также была добавлена амплитудная модуляция. На основе оцененной приведенными методами частот основного тона для каждого сигнала высчитывается параметр *PF1* как мера частотного отклонения в сигнале:

$$PF1 = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{|f(i+1) - f(i)|}{0.5 * [f(i+1) + f(i)]} \times 100 \quad (14)$$

где N – количество частот основного тона, $f(i)$ – частота основного тона в *i*-ом фрагменте сигнала.

Параметр *PF1* используется для определения среднего отклонения первого порядка.

Поскольку все тестовые сигналы моделировались компьютером, теоретические значения частоты основного тона и теоретические *PF1* параметры были известны.

Сравнение полученных параметров пертурбации *PF1* с теоретическими значениями *PF1* изображено на рисунке 11 и с увеличенным масштабом на рисунке 12.

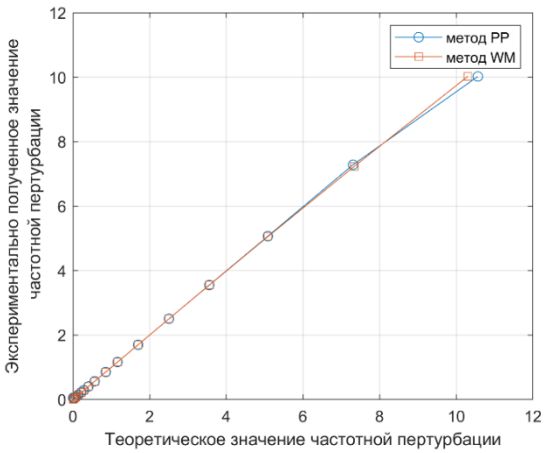

Рисунок 11 – Отношение экспериментально полученной частотной пертурбации к теоретическому значению

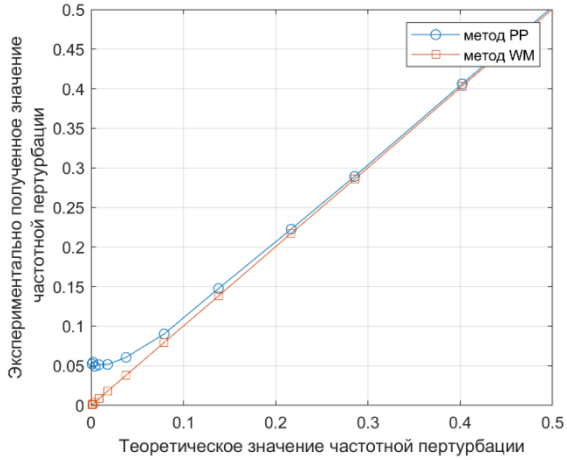

Рисунок 12 – Отношение экспериментально полученной частотной пертурбации к теоретическому значению (увеличенный масштаб)

Таким образом, в работе рассмотрены методы сегментации голосовых сигналов, а именно метод подгонки формы сигнала и метод отбора локальных максимумов. Из полученных графиков видно, что частоты основного тона извлекаются достаточно точно, однако при небольших коэффициентах частотной модуляции лучше использовать метод подгонки формы сигнала. При больших коэффициент частотной модуляции не имеет значения какой метод использовать. Работоспособность методов показана путём MATLAB-моделирования.

Список использованных источников:

1. Comparison of F_0 Extraction Methods for High-Precision Voice Perturbation Measurements / Ingo R. Titze, Haixiang Liang // Journal of Speech and Hearing Research, 1993. – P. 14.

2. A Comparison of Healthy and Disordered Voices Using Multi-Dimensional Voice Program, Praat, and TF32 / Lap-Ching Keung [et al.] // Journal of Voice, 2022 – P. 16.

UDC 616.71

VOICE SIGNAL SEGMENTATION METHODS

Pasternak V.V., Stetsko V.Y.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Vashkevich M.I. – PhD in Technology

Annotation. The paper considers voice signal segmentation methods, namely the Waveform Matching method and the Peak-Picking method. The result of these methods is the extraction of the fundamental frequency of the voice signal. The methods were tested on test signals having a sinusoidal shape and various frequency modulation. To test the performance, the methods were compared in relation of the frequency perturbation parameters to their theoretical values.

Keywords. Voice signal segmentation, fundamental frequency.