

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**  
( Н И У « Б е л Г У » )

ИНСТИТУТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ И ЕСТЕСТВЕННЫХ НАУК  
КАФЕДРА ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ  
СИСТЕМ И ТЕХНОЛОГИЙ

**СУБПОЛОСНАЯ ИДЕНТИФИКАЦИЯ ДИКТОРОВ**

Магистерская диссертация  
обучающегося по направлению подготовки 11.04.02  
Инфокоммуникационные технологии и системы связи,  
магистерская программа «Системы и устройства радиотехники и связи»  
очной формы обучения, группы 07001532  
Кисиленко Анны Владимировны

Научный руководитель  
доктор технических наук,  
профессор, зав. кафедрой  
Информационно -  
телекоммуникационных систем  
и технологий НИУ «БелГУ»  
Жиляков Е.Г.

Рецензент  
канд. техн. наук, доцент кафедры  
Прикладной информатики и  
информационных технологий  
НИУ «БелГУ»  
Черноморец А.А.

БЕЛГОРОД 2017

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1 МЕТОДЫ ИДЕНТИФИКАЦИИ ДИКТОРОВ	7
1.1 Общая схема процедуры идентификации	7
1.2 Основные подходы к построению признакового пространства и их адекватность	9
1.2.1 Спектральные признаки	9
1.2.2 Кепстральные признаки	18
1.2.3 Признаки, извлекаемые на основе субполосного анализа	25
1.3 Решающие правила	34
1.4 Результаты и выводы главы. Задачи исследования	36
ГЛАВА 2 РАЗРАБОТКА МЕТОДА СУБПОЛОСНОЙ ИДЕНТИФИКАЦИИ ДИКТОРОВ	38
2.1 Удаление пауз из речевого сигнала диктора	38
2.2 Сегментация речевых сигналов на однородные отрезки	41
2.3 Решающая функция для субполосной идентификации дикторов	46
2.4 Результаты и выводы главы	48
ГЛАВА 3 ИССЛЕДОВАНИЕ МЕТОДА СУБПОЛОСНОЙ ИДЕНТИФИКАЦИИ ДИКТОРОВ	50
3.1 Реализация этапа обучения	50
3.2 Вычисление порогового значения для метода субполосной идентификации дикторов, где качестве признаков используется распределение энергии сигнала по частотным интервалам	52
3.3 Вычисление порогового значения для метода субполосной идентификации дикторов, где качестве признаков используется распределение долей энергии информационных частотных интервалов	57

3.4 Результаты и выводы главы	58
ЗАКЛЮЧЕНИЕ	60
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	62
ПРИЛОЖЕНИЕ А	70
ПРИЛОЖЕНИЕ Б	78

## **ВВЕДЕНИЕ**

**Актуальность диссертационного исследования.** В связи с развитием автоматических систем человечество находится в поиске удобных средств взаимодействия между техническими средствами и человеком. Одним из актуальных направлений в этой области является разработка систем, предназначенных для идентификации диктора по голосу, которая заключается в определении по образцу голоса, кому из ранее зарегистрированных пользователей принадлежит данный образец. При этом обработке подвергаются речевые сигналы, которые представляют собой результаты регистрации электрических колебаний на выходе микрофона, возбуждаемого акустическими воздействиями.

Проблема идентификации диктора по голосу была поставлена более 40 лет тому назад, и рассматривается в большом количестве работ, среди которых можно отметить работы Л. Рабинера, В.Н. Сорокина, В.Г. Михайлова, А.В. Аграновского, Г.С. Рамишвили и других.

Частным случаем идентификации диктора является идентификация по прецеденту на закрытом множестве дикторов, когда образец голоса диктора, проходящего процедуру идентификации, сравнивается с образцами голосов дикторов, зарегистрированных в системе. При этом количество зарегистрированных дикторов ограничено, а образцом голоса является речевой сигнал, соответствующий произнесённому диктором заранее известному парольному слову.

Такая постановка задачи используется в системах управления роботами («Умный дом», идентификация людей в системах доступа).

Основной проблемой идентификации диктора, как и любой задачи классификации, является выбор признаков, характеризующих особенности речевых сигналов, и обоснование используемых решающих функций.

Применимо к задачам идентификации дикторов, признаки, характеризующие особенности голоса говорящего и лежащие в основе этапа формирования признакового пространства, должны оставаться стабильными на одних и тех же образцах голоса, и при этом резко меняются от диктора к диктору, а также содержат всю необходимую информацию для решения поставленной задачи. Что касается выбранной решающей функции, то она должна сохранять своё значение на идентичных объектах, и быть максимально чувствительной к неидентичности объектов.

Поэтому методы обработки данных, используемые в задачах идентификации диктора для отбора стабильных информативных признаков и удовлетворения условий к решающей функции, должны адекватно отображать свойства анализируемых речевых сигналов, а именно:

- РС являются нестационарными сигналами, т.е. их параметры могут меняться с течением времени;
- порождаемые звуками речи отрезки РС обладают свойством концентрации энергии в достаточно «узких» полосах частотной оси.

Существующие методы идентификации дикторов базируются на спектральном и кепстральном подходах для поиска стабильных информативных признаков. Использование признаков, вычисляемых на основе подобных подходов, не всегда удовлетворяет предъявляемым требованиям или не является адекватным с точки зрения учёта свойств речевых сигналов. Поэтому поиск новых признаков, характеризующих особенности голоса говорящего, является актуальным.

В данной работе для поиска подобных признаков предлагается использование субполосного анализа, когда свойства речевых сигналов соотносятся с некоторым разбиением оси частот на интервалы конечной ширины.

**Целью данной работы является** совершенствование методов и алгоритмов идентификации дикторов на основе разработки субполосного метода идентификации дикторов по прецеденту.

Для достижения поставленной цели сформулированы и решены следующие задачи:

1. Анализ основных подходов, используемых в задачах идентификации дикторов;
2. Разработка решающей функции для идентификации диктора на основе субполосного подхода;
3. Разработка и исследование субполосного метода и алгоритма идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение частей энергии по частотным интервалам;
4. Разработка и исследование субполосного метода и алгоритма идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение долей энергии информационных частотных интервалов;
5. Проведение вычислительных экспериментов.

**Объект исследований:** речевые сигналы.

**Предмет исследования:** методы идентификации дикторов.

**Методы исследований.** В работе использованы методы субполосного анализа/синтеза, статистического анализа, цифровой обработки речевых данных, линейной алгебры и вычислительных экспериментов.

**Объем и структура работы.** Диссертация состоит из Введения, трёх глав, Заключение и Приложений. Работа изложена на 87 страницах машинописного текста, включая 69 рисунков, 2 таблицы и список литературных источников из 71 наименования.

# ГЛАВА 1 МЕТОДЫ ИДЕНТИФИКАЦИИ ДИКТОРОВ

## 1.1 Общая схема процедуры идентификации

Задача идентификации диктора заключается в определении по образцу голоса, кому из ранее зарегистрированных пользователей принадлежит данный образец.

Процедура идентификации происходит следующим образом: сначала проходит этап обучения, в результате которого происходит отбор множества признаков, на основе которых происходит сравнение записей дикторов, и обоснование решающей функции и критической области, которая используется для принятия решения.

1. Отбор множества признаков представляет собой отбор некоторых параметров, которые отражают особенности голоса говорящего и получаются на основе обработки речевых сигналов. Т.е. каждый из интересующих дикторов регистрируется в системе идентификации посредством предоставления образца своего голоса – отрезка речевого сигнала, соответствующего произнесённому диктором слову или фразе. На основе отрезков речевых данных, которые являются цифровым представлением речевых сигналов, производится обучение системы идентификации, а именно из каждого отрезка извлекаются стабильные информативные признаки. Под стабильными признаками понимаются такие, которые остаются стабильными на одних и тех же образцах голоса, и при этом резко меняются от диктора к диктору. Под информативными признаками будем понимать те, которые содержат всю необходимую информацию для решения поставленной задачи. Этап поиска и выделения подобных признаков называется формирование признакового пространства. Эти признаки хранятся в памяти вычислительного устройства, тем самым формируя базу данных дикторов.

После этапа формирования базы, человек, который хочет пройти процедуру идентификации, вновь предоставляет образец своего голоса, из которого извлекаются признаки. После чего признаки, полученные при прохождении процедуры идентификации, сопоставляются с признаками, хранимыми в базе. Сопоставление признаков осуществляется на основе анализа решающей функции и сравнения с пороговыми значениями, которые определяют вероятности ошибочного принятия решения (критическая область). Под решающей функцией понимается способ обработки данных и их интерпретации с позиции решаемой задачи.

Для того, чтобы построить решающую функцию, формулируется изначальное предположение (Начальная гипотеза  $H_0$  –признаки, полученные при прохождении процедуры идентификации, и признаки, хранимые в базе, принадлежат разным дикторам). После чего для гипотезы строится критическая область, когда гипотеза должна отвергаться. Гипотезу нужно сформулировать таким образом, чтобы вероятность попадания значения решающей функции в критическую область была достаточно мала. Обеспечить это можно, только зная вероятностные свойства исследуемого объекта, т.е. проведя и проанализировав ряд предварительных экспериментов, тем самым формируя обучающую выборку.

Применимо к задаче идентификации решающая функция должна удовлетворять следующим требованиям:

1. На идентичных объектах решающая функция должна сохранять своё значение (т.е. быть инвариантной);
2. Решающая функция должна быть максимально чувствительна к неидентичности объектов.

Данные требования можно свести к вариационной задаче, а именно, что необходимо минимизировать вероятность ошибки II рода при фиксированном значении вероятности ошибки I рода.

Результатом процедуры идентификации является имя человека из общей совокупности дикторов, зарегистрированных в системе, речевой сигнал которого наиболее вероятно совпадает с образцом голоса идентифицируемого диктора.

Таким образом, задача идентификации диктора включает в себя:

- формирование признакового пространства;
- построение решающей функции.

## **1.2 Основные подходы к построению признакового пространства и их адекватность**

На этапе формирования признакового пространства устойчивые информативные признаки выделяются из речевых сигналов пользователей. Поэтому, оценивая адекватность подходов к построению признакового пространства, мы будем говорить о том, насколько они отражают основные свойства речевых сигналов, а именно:

1. РС являются нестационарными сигналами, т.е. их параметры могут меняться с течением времени;
2. Порождаемые звуками речи отрезки РС обладают свойством концентрации энергии в достаточно «узких» полосах частотной оси.

Рассмотрим основные подходы к построению признакового пространства и оценим их адекватность с позиции учёта оговорённых свойств.

### **1.2.1 Спектральные признаки**

Речевой сигнал, как и другие виды сложных сигналов произвольной формы, при их изучении и анализе, удобно представлять в виде некоторой

комбинации более простых сигналов известной формы. Такой приём – разложение произвольной функции по различным системам детерминированных базисных функций широко используется в цифровой обработке сигналов и называется обобщённым спектральным анализом [21].

Так, если в качестве базисных функций использовать ортогональную на отрезке  $[a, b]$  бесконечную систему непрерывных функций вида

$$\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x), \dots \quad (1.1)$$

где выполняются условия:

$$\int_a^b \varphi_n(x) \varphi_m(x) dx = 0 \quad (1.2)$$

$$\int_a^b \varphi_n^2(x) dx \neq 0 \quad (1.3)$$

то произвольная кусочно – непрерывная функция  $f(x)$ , для которой выполняется условие

$$\int |f(x)|^2 dx < \infty \quad (1.4)$$

может быть представлена в виде бесконечной суммы ряда:

$$f(x) = c_0 \varphi_0(x) + c_1 \varphi_1(x) + \dots + c_n \varphi_n(x) + \dots \quad (1.5)$$

где  $\varphi_n(x)$  – система ортогональных непрерывных функций,  $c_n$  – коэффициенты ряда.

В случае если коэффициенты ряда  $c_n$  определяются как в [14], а именно:

$$c_n = \frac{1}{\|\varphi_n\|^2} \int_a^b f(x) \varphi_n(x) dx \quad (1.6)$$

то ряд (1.5) называется обобщённым рядом Фурье [14]. Вся информация о  $f(x)$  при заданной системе функций  $\varphi_n(x)$  содержится в наборе коэффициентов  $c_i$  (спектре).

Обобщённый ряд Фурье обладает важным свойством: при заданной системе функций  $\varphi_n(x)$  и фиксированном числе слагаемых  $c_i$  (т.е.  $i = \overline{0, N}$ ), он обеспечивает наилучшую аппроксимацию функции  $f(x)$ , что подтверждается неравенством Бесселя:

$$\sum_{i=0}^N c_i^2 \|\varphi_i\|^2 \leq \|f\|^2 \quad (1.7)$$

Иными словами, если бесконечная сумма ряда (1.5) позволяет восстановить  $f(x)$  по набору коэффициентов  $c_i$  абсолютно точно, то при ограничении числа членов ряда никакой другой способ разложения не может дать лучшего приближения суммы (1.5) к функции  $f(x)$  [21].

При этом количество слагаемых ряда  $N$  выбирается исходя из соображений выполнения условия:

$$\int_a^b \left[ f(x) - \sum_{i=0}^N c_i \varphi_i(x) \right]^2 dx \leq \delta^2 \int_a^b f^2(x) dx \quad (1.8)$$

где  $\delta^2$  - желаемая (требуемая) относительная погрешность.

Для непериодических сигналов конечной длительности  $f(t)$ , какими и являются речевые сигналы, используется формула разложения, при которой дискретность или -шаг вычисления спектра стремится по величине к нулю и дискретный ряд Фурье переходит в интеграл Фурье или преобразование (трансформанту) Фурье:

$$f(x) = \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} dt \quad (1.9)$$

где  $F(\omega)$  - частотное представление сигнала  $f(t)$ .

И обратное преобразование Фурье:

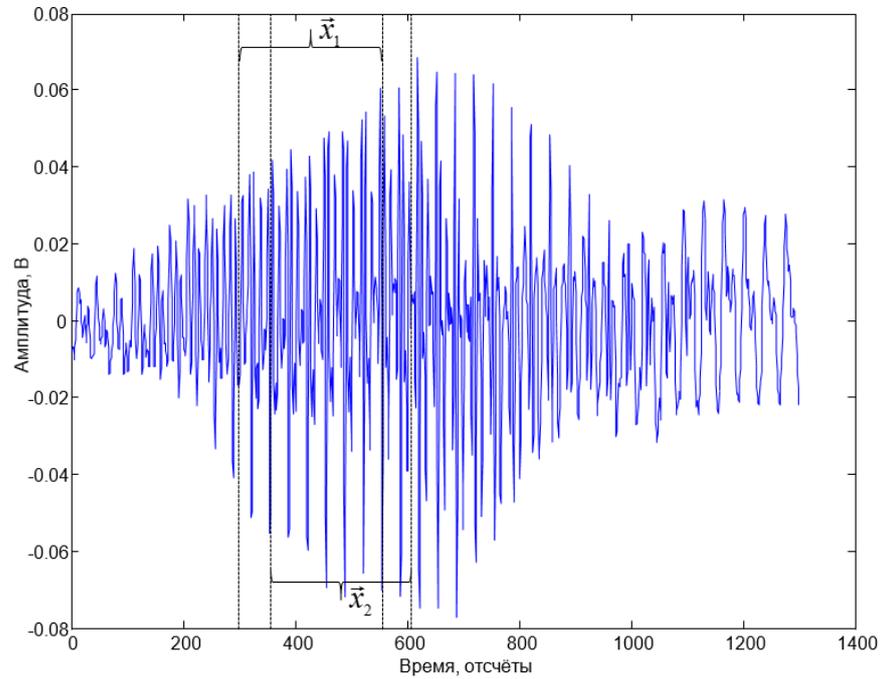
$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (1.10)$$

Соотношения (1.9) и (1.10) являются эквивалентными, поскольку могут быть получены друг из друга линейным преобразованием.

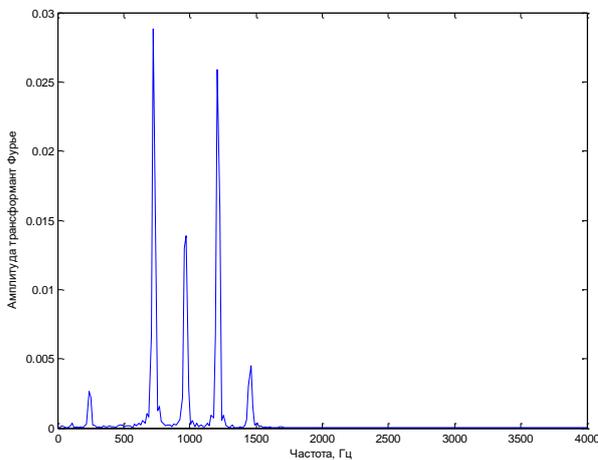
Значение  $F(\omega)$  используется как набор признаков, характеризующий особенности голоса диктора. Для того, чтобы оценить, насколько подобный подход отвечает условиям стабильности и информативности извлекаемых признаков, посмотрим, как ведут себя трансформанты Фурье при смене дикторов, смене парольного слова. При этом обработке подвергаются отрезки речевых сигналов, в пределах которых характеристики речевого сигнала остаются неизменными. Такого рода результат можно получить посредством кратковременного анализа сигнала: считается, что энергия (а следовательно и частотные характеристики) сигнала является постоянной характеристикой на интервале 10-30 мс.

На рисунке 1.1 представлен речевой сигнал, соответствующий произнесению фонемы «а». В этом сигнале были выделены два отрезка  $\vec{x}_1$  и  $\vec{x}_2$ , каждый длиной 256 отсчётов при частоте дискретизации  $f_0 = 8000$  Гц (т.е. длительность отрезков составляет 32 мс). При этом отрезки выбраны со сдвижкой в 50 отсчётов.

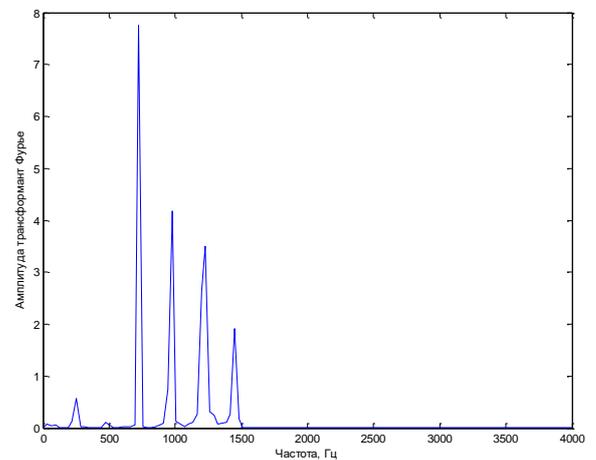
На рисунках 1.2.А – 1.2.Б представлены трансформанты Фурье отрезков  $\vec{x}_1$  и  $\vec{x}_2$  соответственно.



**Рисунок 1.1 – РС, соответствующий произнесению фонемы «а», выбор отрезков анализа  $x_1$  и  $x_2$**



**А**

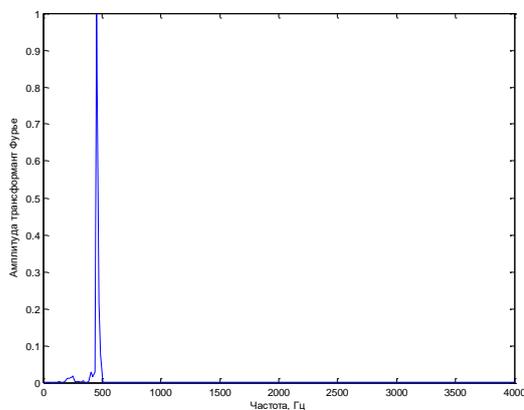


**Б**

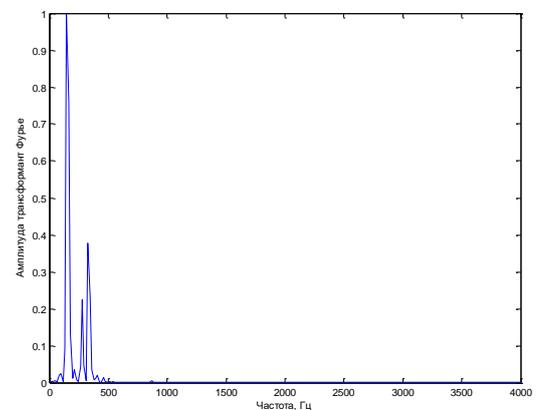
**Рисунок 1.2 – Трансформанты Фурье отрезков сигналов  $x_1$  (А) и  $x_2$  (Б)**

На основании рисунков 1.2.А – 1.2.Б можно сделать вывод, что распределение энергии меняется в зависимости от отрезка анализа, т.е. нестабильно.

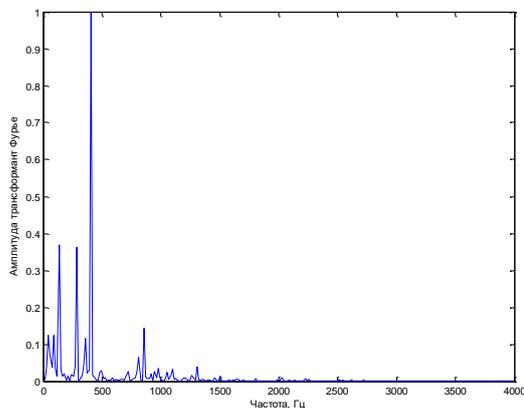
Для того, чтобы продемонстрировать, что признаки, получаемые на основе частотных представлений – чаще всего это трансформанта Фурье – отвечают требованию резкой смены от диктора к диктору, на рисунках 1.3 - 1.4 представлены спектры сигналов, соответствующие произнесению фонемы «т» из слова «четыре» (рисунок 1.3) и фонемы «о» из слова «восемь» (рисунок 1.4) разными дикторами.



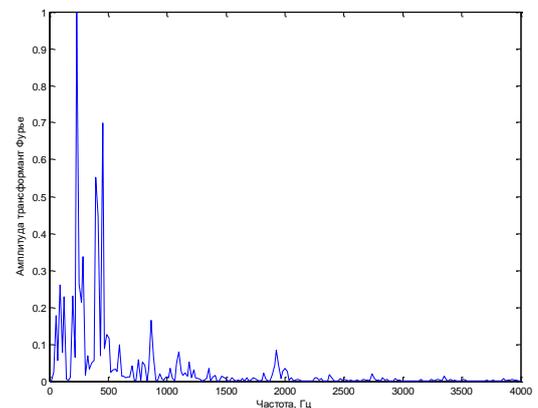
А



Б

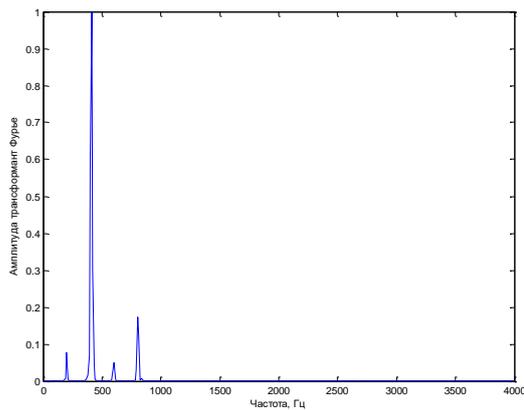


В

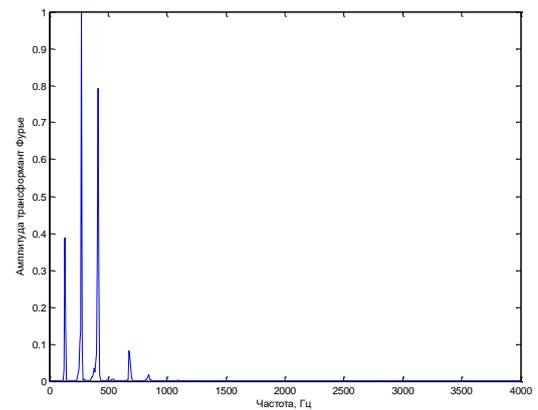


Г

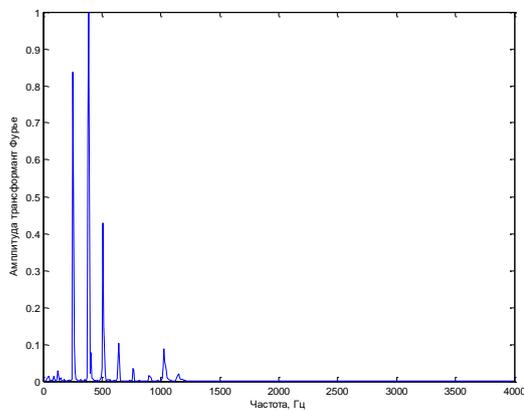
**Рисунок 1.3 – Трансформанта Фурье сигналов, соответствующих произнесению фонемы «т» из слова «четыре» Диктором 1(А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**



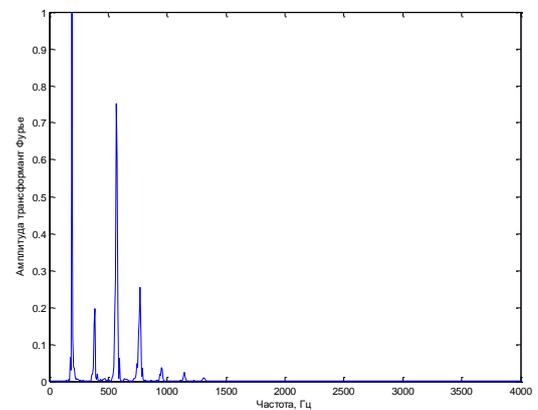
А



Б



В



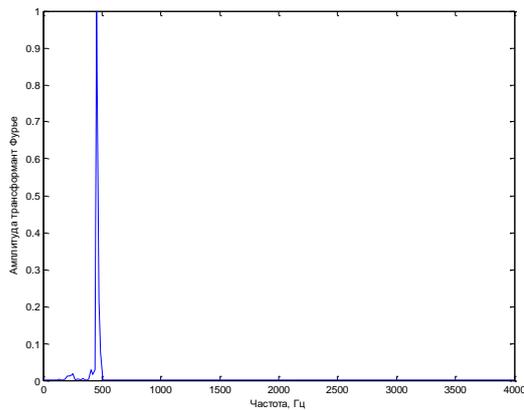
Г

**Рисунок 1.4 – Трансформанта Фурье сигналов, соответствующих произнесению фонемы «о» из слова «восемь» Диктором 1(А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**

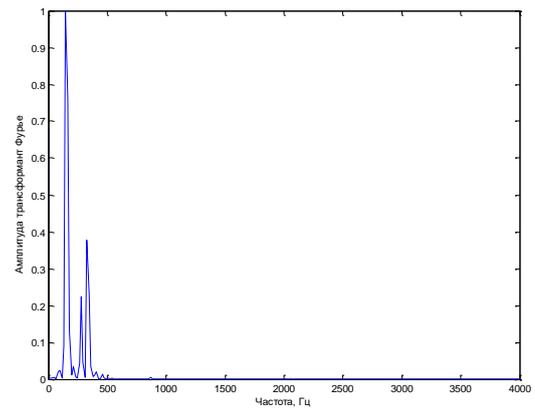
На основании рисунков 1.3 – 1.4 можно сделать вывод, что формы спектров для одних и тех же фонем при произнесении их разными дикторами, меняются. Поэтому трансформанта Фурье как признак, используемый для процедуры идентификации, отвечает условию резкой смены от диктора к диктору.

Но использование трансформанты Фурье как признака, характеризующего особенности голоса диктора, не отвечает условию стабильности признаков на одном и том же дикторе. Для того, чтобы

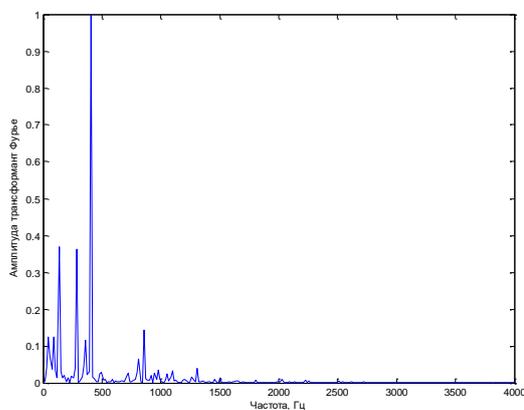
продемонстрировать, что признаки, получаемые на основе спектрального анализа, не отвечают условию стабильности, на рисунках 1.5 – 1.6 представлены спектры сигналов, соответствующие произнесению фонемы «т» из слова «четыре» (рисунок 1.5) и фонемы «о» из слова «восемь» (рисунок 1.6), соответствующие одному диктору.



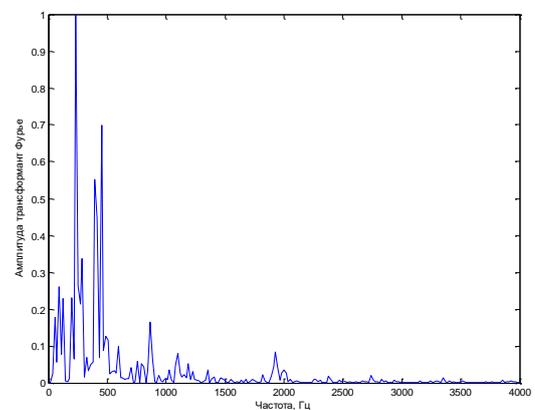
А



Б

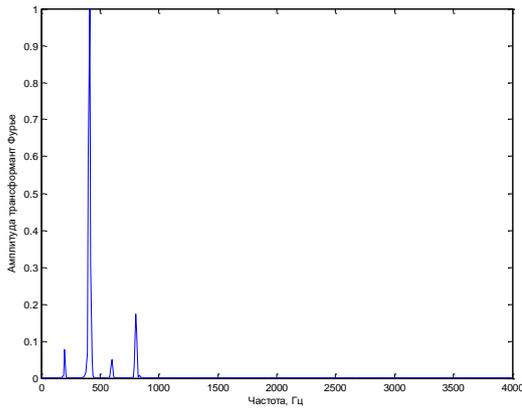


В

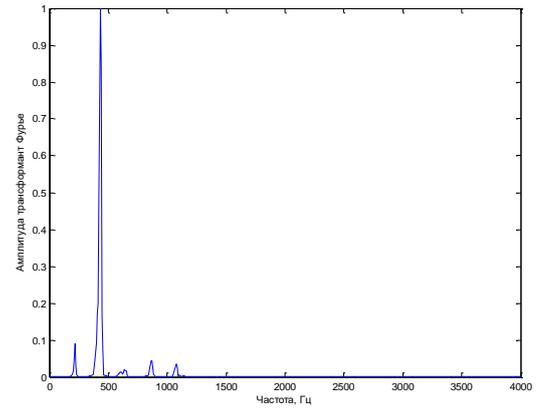


Г

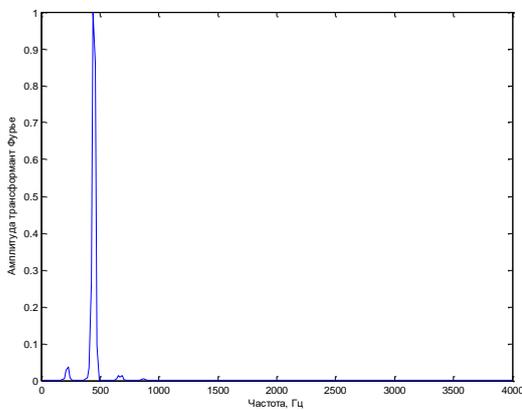
**Рисунок 1.5 – Трансформанта Фурье сигналов, соответствующих произнесению фонемы «т» из слова «четыре» Диктором 1 в четырёх различных случаях**



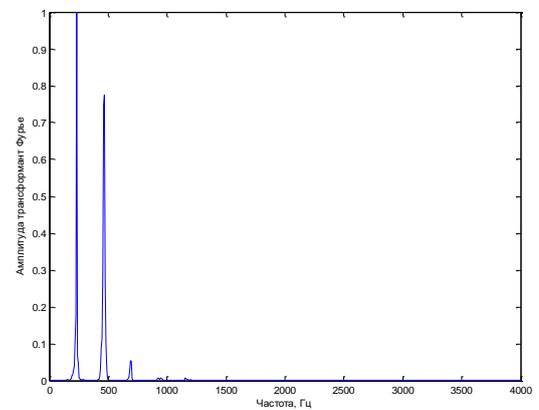
А



Б



В



Г

**Рисунок 1.6 – Трансформанта Фурье сигналов, соответствующих произнесению фонемы «о» из слова «восемь» Диктором 1 в четырёх различных случаях**

Таким образом, спектральный анализ является адекватным описанием речевого сигнала, поскольку позволяет учитывать свойство сосредоточенности энергии РС, но признаки, получаемые на основе такого подхода, не являются стабильными.

### 1.2.2 Кепстральные признаки

Кепстральный анализ базируется на модели генерации речевых сигналов.

Предполагается, что речевой сигнал  $s(t)$  может быть представлен в виде:

$$s(t) = \int_{-\infty}^{\infty} k(t - \tau)u(\tau)d\tau, \quad (1.11)$$

где  $k(t)$  – импульсная характеристика;  
 $u(t)$  – входное воздействие.

Для того, чтобы отделить влияние полости от входного воздействия, воспользуемся теоремой о свёртке:

$$S(\omega) = K(\omega) \cdot U(\omega), \quad (1.12)$$

где  $S(\omega)$  – спектр речевого сигнала  $s(t)$ ;  
 $K(\omega)$  – спектр импульсной характеристики;  
 $U(\omega)$  – спектр входного воздействия.

При этом допускается, что диапазоны частот, в которых сосредоточены энергии  $K(\omega)$  и  $U(\omega)$  различны и не пересекаются, т.е. ограничены частотой  $\Omega$ , а импульсная характеристика является низкочастотной.

$$\begin{aligned} k(t) &= \int_{\omega \in \Omega} K(\omega)e^{i\omega t} d\omega \\ u(t) &= \int_{\omega \notin \Omega} U(\omega)e^{i\omega t} d\omega \end{aligned} \quad (1.13)$$

Тогда произведение спектров двух сигналов из выражения (1.12) можно представить в виде суммы логарифмов квадратов модулей спектров:

$$|S(\omega)|^2 = |K(\omega)|^2 \cdot |U(\omega)|^2, \quad (1.14)$$

$$\ln|S(\omega)|^2 = \ln|K(\omega)|^2 + \ln|U(\omega)|^2 \quad (1.15)$$

Принцип, заключённый в представлении (1.12) через (1.15), называется гомоморфной фильтрацией.

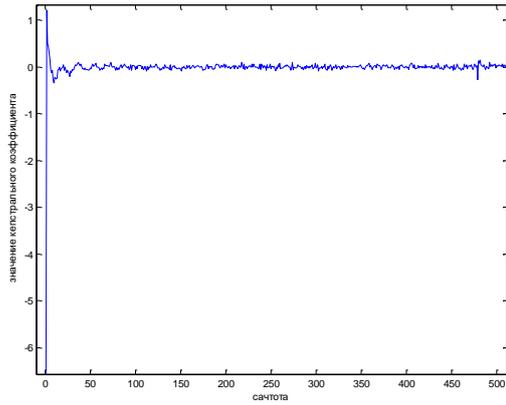
При этом кепстры, которые как раз и используются для формирования признакового пространства, определяются как обратное преобразование от логарифма квадрата спектра:

$$C_s(q) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln|S(\omega)|^2 e^{i\omega q} d\omega, \quad (1.16)$$

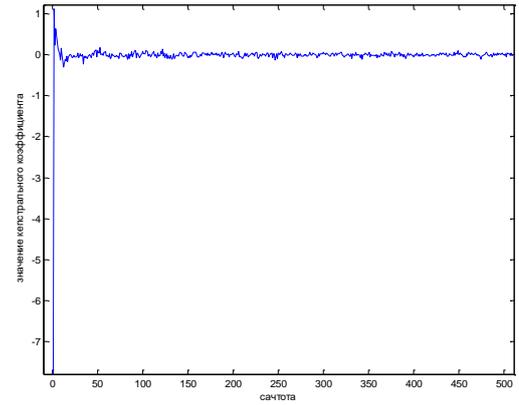
где  $C_s(q)$  – кепстр мощности;

$q$  – сачтота, отсчёты.

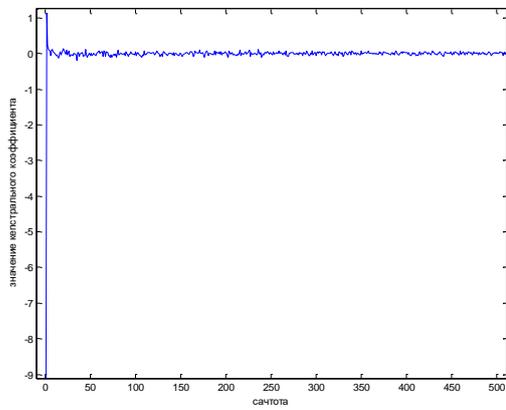
Для того, чтобы продемонстрировать, насколько использование кепстра в качестве вектора признаков отвечает условию стабильности, вычислим значения кепстральных коэффициентов для отрезков речевых сигналов, соответствующих фонеме «т» в слове «четыре» и фонеме «о» в слове «восемь», произнесённые разными дикторами.



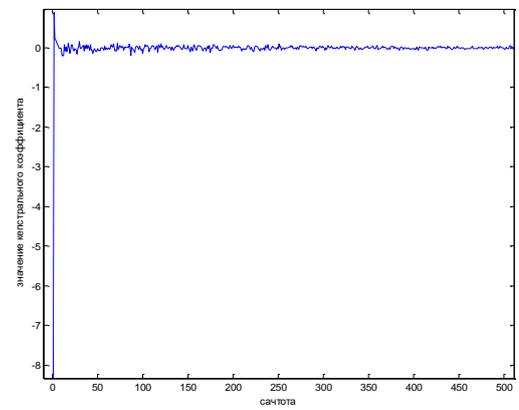
А



Б

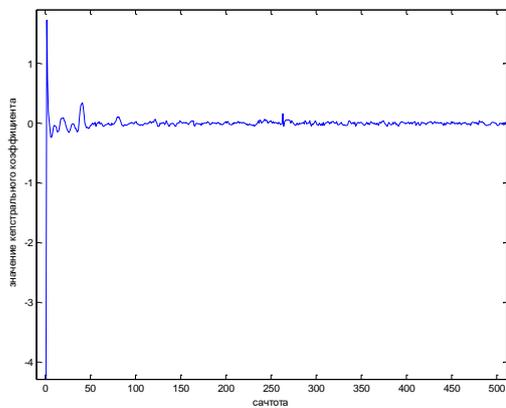


В

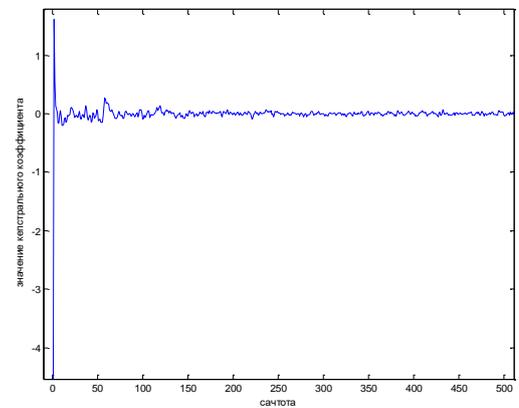


Г

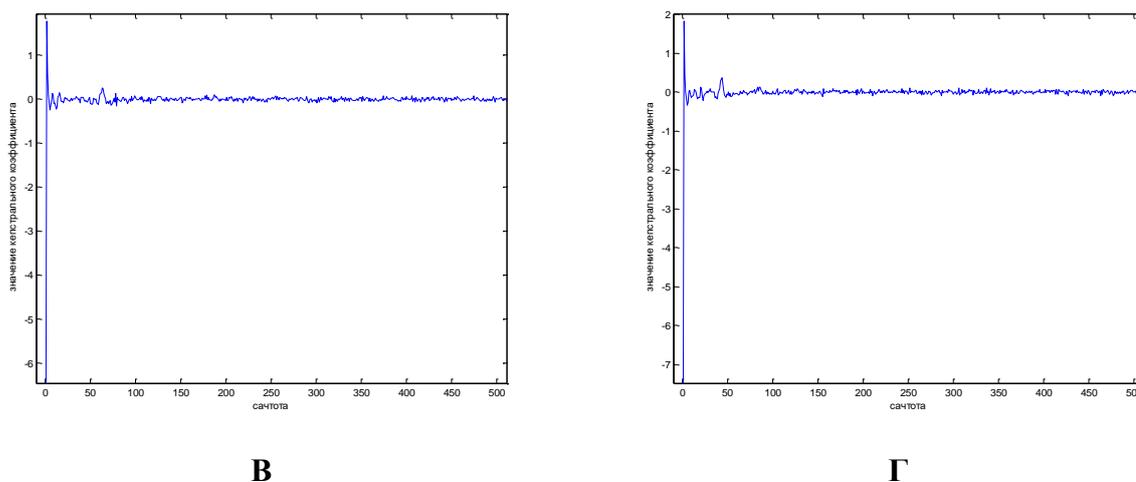
**Рисунок 1.7 – Кепстры речевых сигналов, соответствующих произнесению фонемы «т» из слова «четыре» Диктором 1(А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**



А

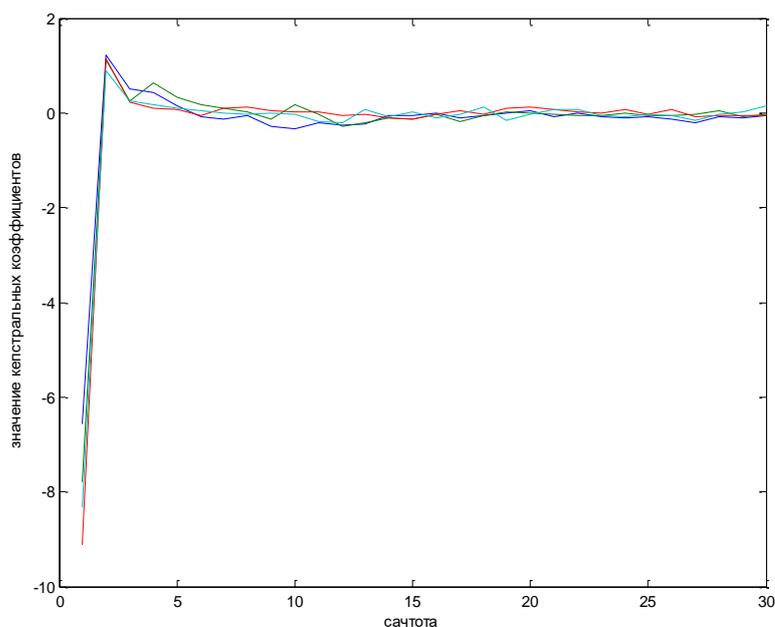


Б

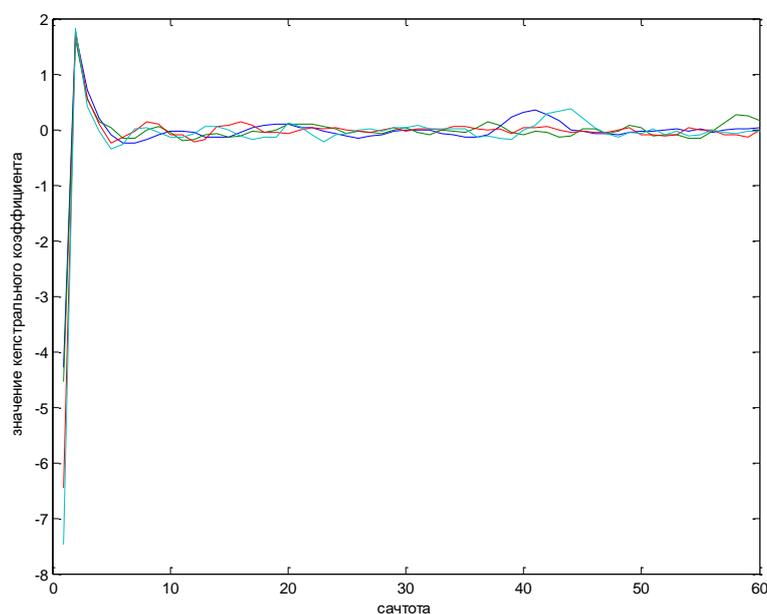


**Рисунок 1.8 – Кепстры речевых сигналов, соответствующих произнесению фонемы «о» из слова «восемь» Диктором 1 (А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**

Если рассмотреть одновременно графики, представленные на рисунке 1.7.А – 1.8.Г, то можно сделать вывод, что значения коэффициентов кепстра для разных дикторов отличаются не настолько, чтобы их можно было разделить (см. рисунок 1.9, 1.10).



**Рисунок 1.9 – 1-30 коэффициенты кепстра речевых сигналов, соответствующих произнесению фонемы «т» из слова «четыре» Дикторами 1 – 4**

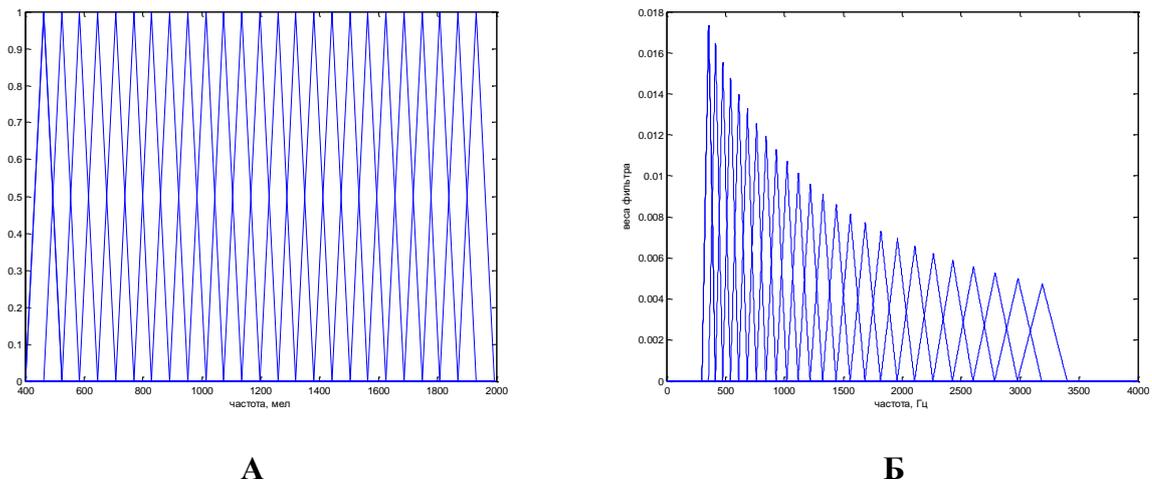


**Рисунок 1.10 – 1-60 коэффициенты кепстра речевых сигналов, соответствующих произнесению фонемы «о» из слова «восемь» Дикторами 1 – 4**

Дальнейшим витком развития кепстрального подхода в применении к задачам идентификации диктора стало вычисление мел – кепстральных частотных коэффициентов, использование которых предположительно позволит учитывать свойства слуха.

Мел — психофизическая единица высоты звука. Считается, что учёт психофизических свойств уха позволит улучшить эффективность алгоритмов идентификации диктора.

При вычислении мел – кепстральных частотных коэффициентов используется гребёнка треугольных полосовых фильтров. При этом граничные частоты фильтров определяются путем нелинейного преобразования шкалы частот из шкалы Герц в шкалу Мел (см. рисунок 1.11). Алгоритм вычисления граничных частот фильтра и весовых коэффициентов фильтра можно найти в работах [18, 33, 44].



**Рисунок 1.11 – Вид полосовых фильтров в шкале Мел (А) и их вид в шкале Гц (Б)**

Тогда мел – кепстральные частотные коэффициенты вычисляются в соответствии с (1.17):

$$c_i = \sum_{m=1}^{N_{FB}} \ln(e_m) \cdot \cos\left(\frac{\pi i(m-0,5)}{N_{FB}}\right), \quad i = \overline{1, N_{MFCC}}, \quad (1.17)$$

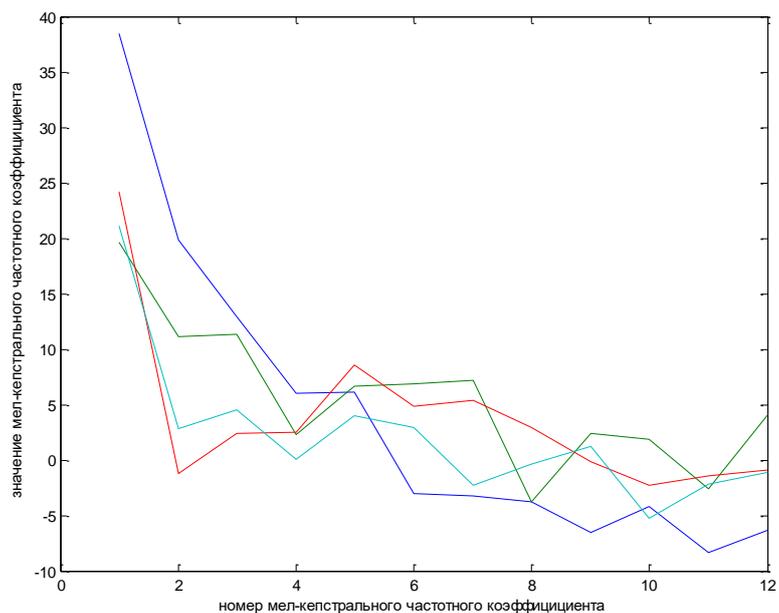
где  $c_i$  - значение мел – кепстрального частотного коэффициента;

$e_m$  – значение энергии с весом на выходе  $m$  – ого полосового треугольного фильтра;

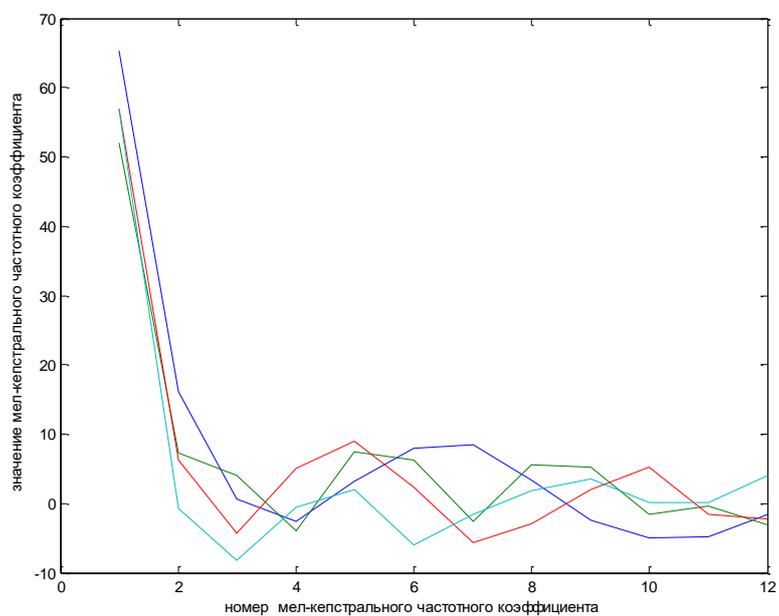
$N_{FB}$  – количество полосовых треугольных фильтров;

$N_{MFCC}$  – количество мел - кепстральных коэффициентов (на практике выбирают порядка 12 [19].).

На рисунках 1.12-1.13 показаны графики значений мел-кепстральных частотных коэффициентов для фонем «т» и «о», произнесённых разными дикторами.



**Рисунок 1.12 – Графики значений мел-кепстральных частотных коэффициентов для речевых сигналов, соответствующих произнесению фонемы «т» из слова «четыре» Дикторами 1 – 4**



**Рисунок 1.13 – Графики значений мел-кепстральных частотных коэффициентов для речевых сигналов, соответствующих произнесению фонемы «о» из слова «восемь» Дикторами 1 – 4**

При использовании мел-кепстральных частотных коэффициентов в качестве стабильных информативных признаков вместо кепстра, нельзя утверждать, что они резко меняются при смене диктора (рисунки 1.12-1.13).

Метод кепстрального представления, т.е. обратного Фурье-преобразования от логарифмического амплитудного спектра в масштабе частот мел, был разработан в результате длительных поисков такого описания речевого сигнала, которое было бы устойчиво к индивидуальным особенностям дикторов. Поэтому довольно странно было бы его использовать для решения противоположной задачи – поиска описания, подчеркивающего различия в голосах дикторов. Однако большинство работ либо игнорирует этот факт, либо занято поиском наиболее эффективных методов принятия решения в заданном пространстве кепстральных коэффициентов [55].

### 1.2.3 Признаки, извлекаемые на основе субполосного анализа

Под субполосной обработкой понимается анализ или синтез сигналов с точки зрения некоторого разбиения частотной полосы на ряд интервалов  $R$  (субполос):

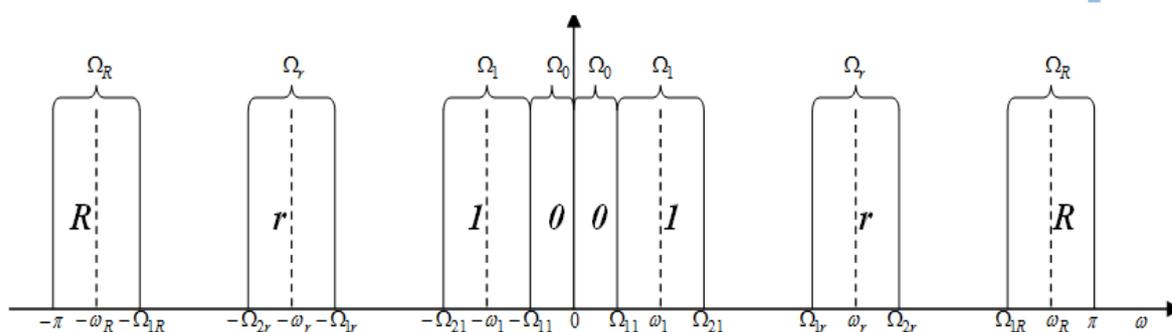


Рисунок 1.14 – Разбиение оси частот на субполосы

На рисунке 1.14  $\Omega_r$  - составной интервал оси нормированных частот:

$$\Omega_r = [-\Omega_{1r}, -\Omega_{2r}) \cup [\Omega_{1r}, \Omega_{2r}), \quad 0 \leq \Omega_{1r} < \Omega_{2r} \leq \pi, \quad (1.18)$$

где  $r$  - номер  $\Omega_r$  - того частотного интервала,  $r = 1, 2, \dots, R$ ;

$\omega_r$  - середина  $\Omega_r$  - того частотного интервала;

$(\Omega_{1r}, \Omega_{2r})$  - границы  $\Omega_r$  - того частотного интервала.

$\Omega_0$  - частотный интервал, ширина которого в 2 раза меньше ширины  $\Omega_r$  - того частотного интервала.

При этом ширины частотных интервалов определяются в соответствии с (1.19) – (1.20), учитывая, что разбиение частотной оси происходит в соответствии с рисунком 1.14:

$$\Delta\omega_0 = \pi/(2R+1), \quad (1.19)$$

где  $\Delta\omega_0$  - ширина  $\Omega_0$  - ого частотного интервала.

$$\Delta\omega = 2\Delta\omega_0, \quad (1.20)$$

где  $\Delta\omega$  - ширина  $\Omega_r$  - ого частотного интервала,  $r = 1, 2, \dots, R$ .

Исходя из того, что частотные интервалы  $r = 1, 2, \dots, R$  равной ширины, то центральную частоту можно получить из соотношения:

$$\omega_r = \Delta\omega_0 \cdot 2r \quad (1.21)$$

Целесообразность такого подхода к обработке речевых сигналов определяется тем, что подавляющие доли их энергии сосредоточены в достаточно малой доле частотной полосы, что является основным свойством, присущим речевым сигналам.

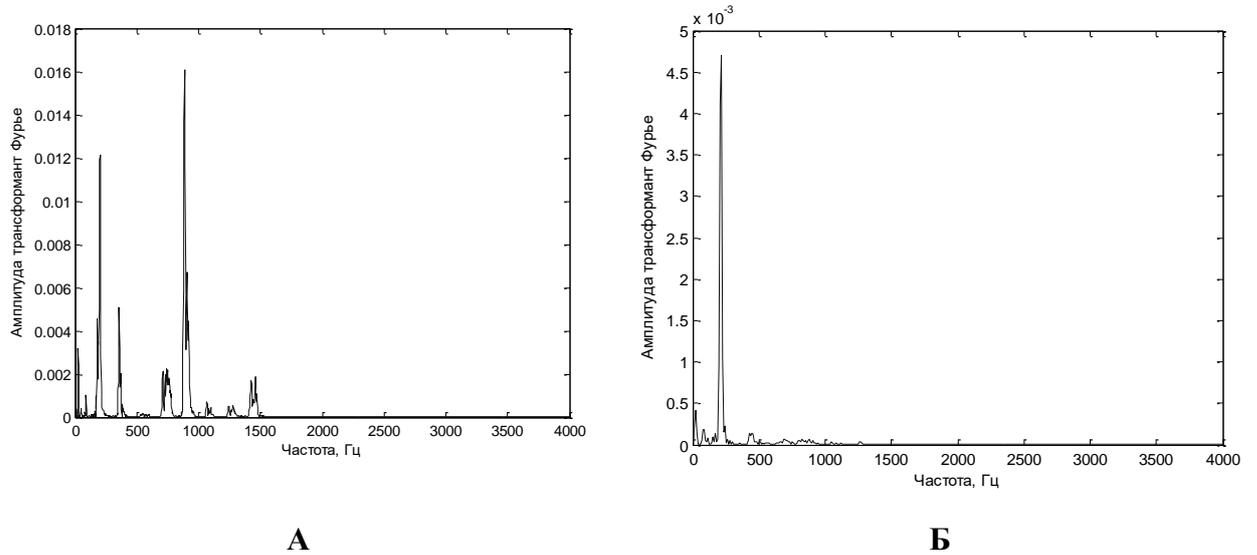


Рисунок 1.15 – Распределение энергии для звуков "а" (А) и "в" (Б) ( $f_d = 8$  кГц)

Основными инструментами субполосной обработки является соотношение для вычисления части энергии  $P_r(\vec{x})$ , сосредоточенной в  $\Omega_r$  - ом частотном интервале сигнала  $\vec{x}$ :

$$P_r(\vec{x}) = \frac{1}{2\pi} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega, \quad r = 1, 2, \dots, R, \quad (1.22)$$

где  $X(\omega)$  - трансформанта (спектр) анализируемого отрезка сигнала  $\vec{x}$ ;

$\Omega_r$  - составной интервал оси нормированных частот.

В качестве стабильных информативных признаков может использоваться распределение энергии отрезка сигнала по интервалам  $P_r(\vec{x})$ .

Используя полученное в монографии [19] соотношение для определения части энергии в субполосе с использованием набора субполосных матриц  $A_r$  (1.25), (1.22) можно привести к виду ф.(1.23):

$$P_r(\bar{x}) = \bar{x}A_r\bar{x}^T, \quad r = 1, 2, \dots, R \quad (1.24)$$

где  $A_r = \{a_{ik}^r\}$  – субполосная матрица для  $\Omega_r$  - того частотного интервала:

$$A_r = \{a_{ik}^r\} = 2 \frac{\sin[\Delta\omega_0(i-k)]}{\pi(i-k)} \cdot \cos(\omega_r(i-k)) = 2a_{ik}^0 \cdot \cos(\omega_r(i-k)), \quad (1.25)$$

где  $\{a_{ik}^r\}$  – элемент симметричной и положительно определенной субполосной матрицы.

$i$  - позиция элемента в строке матрицы,  $i = 1, 2, \dots, N$  ;

$k$  - позиция элемента в столбце матрицы,  $k = 1, 2, \dots, N$  ;

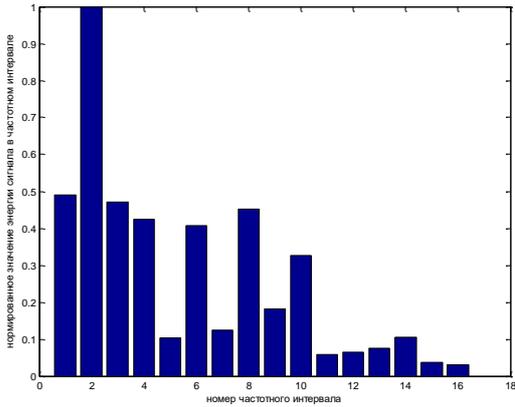
$r$  – индекс, который обозначает номер частотного интервала,  $r = 1, 2, \dots, R$  ;

$\omega_r$  – центральная частота  $\Omega_r$  - того частотного интервала  $|\omega_r| \in [0, \pi]$  ;

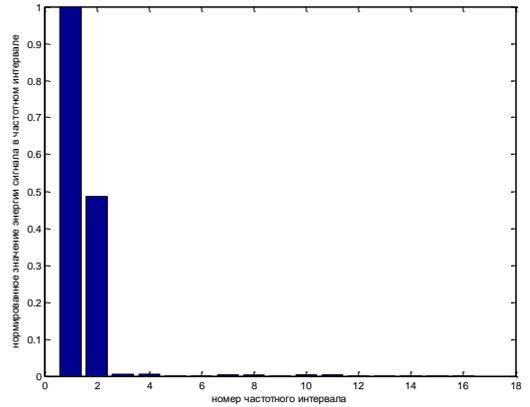
$a_{ik}^0$  – постоянная составляющая:

$$a_{ik}^0 = \frac{\sin[\Delta\omega_0(i-k)]}{\pi(i-k)} \quad (1.26)$$

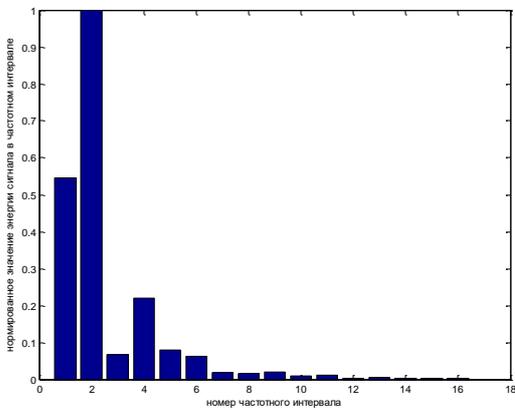
Для того, чтобы продемонстрировать, что признаки, получаемые на основе субполосного анализа – распределение энергии сигнала по частотным интервалам – отвечают требованию резкой смены от диктора к диктору, на рисунках 1.16 – 1.17 представлены распределения энергии по частотным интервалам сигналов, соответствующих произнесению фонемы «т» из слова «четыре» (рисунок 1.16) и фонемы «о» из слова «восемь» (рисунок 1.17) разными дикторами.



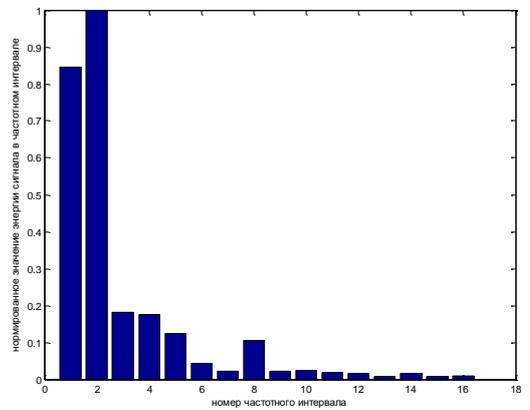
**А**



**Б**

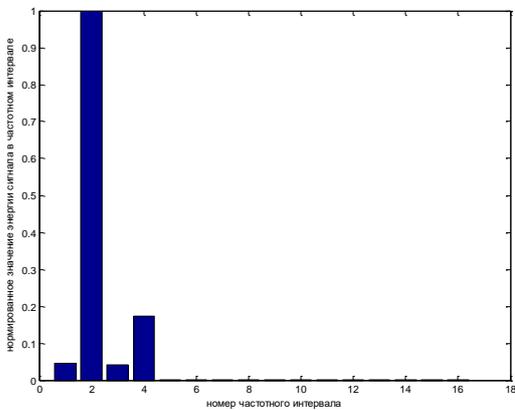


**В**

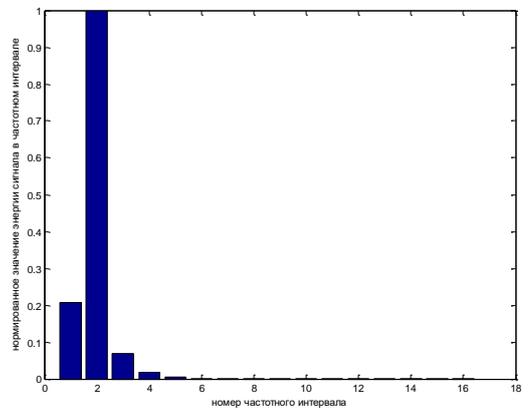


**Г**

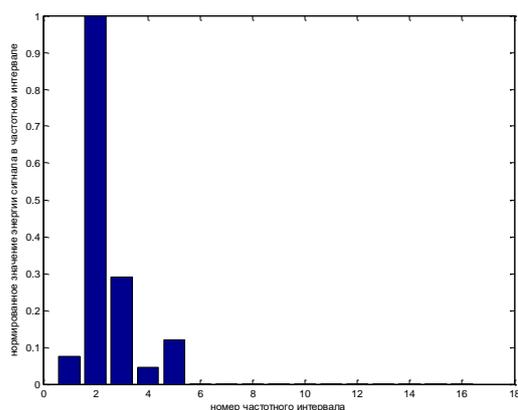
**Рисунок 1.16 – Распределение энергии по частотным интервалам, где в качестве отрезка сигнала выступает фонема «т» из слова «четыре» Диктором 1(А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**



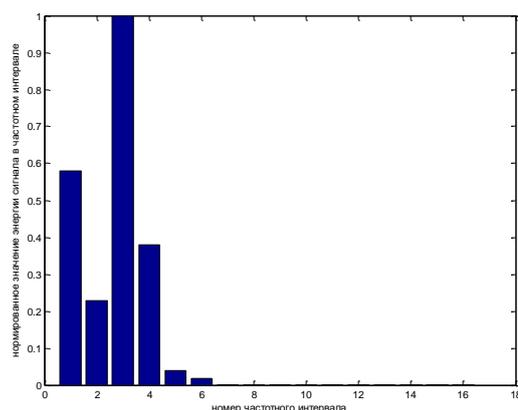
**А**



**Б**



В

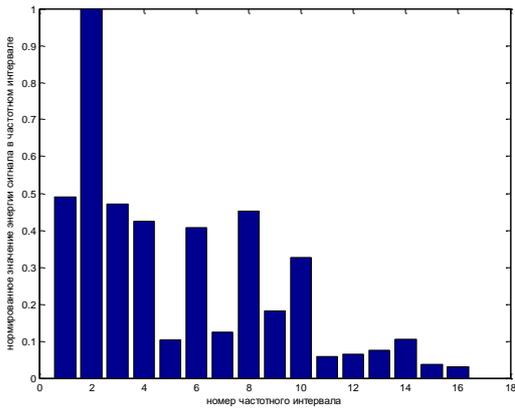


Г

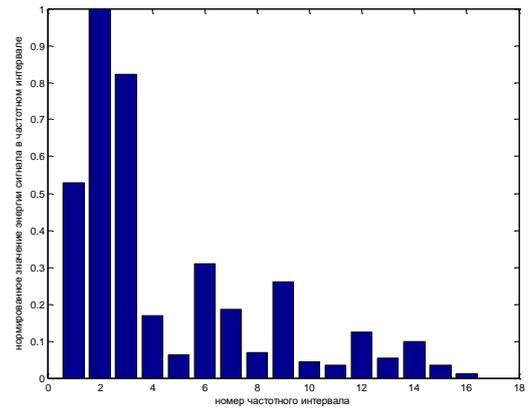
**Рисунок 1.17 – Распределение энергии по частотным интервалам, где в качестве отрезка сигнала выступает фонема «о» из слова «восемь», по частотным интервалам Диктором 1 (А), Диктором 2 (Б), Диктором 3 (В) и Диктором 4 (Г)**

На основании рисунков 1.16 – 1.17 можно сделать вывод, что использование распределения энергии по частотным интервалам как признака для формирования признакового пространства, отвечает условию резкой смены от диктора к диктору.

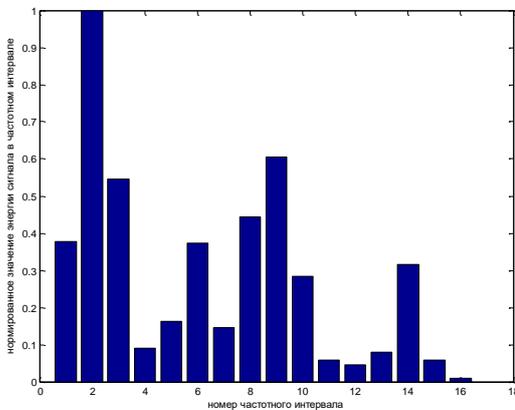
Т.к. признаки, используемые в процедурах идентификации диктора, должны также быть стабильны в одних и тех же условиях – то есть на одних и тех же фонемах, произнесённых одним и тем диктором, – то на рисунках 1.18 – 1.19 показано, как ведёт себя распределение энергии по частотным интервалам в условиях данной задачи.



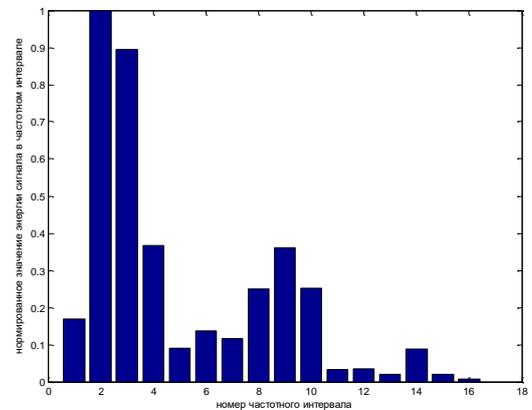
А



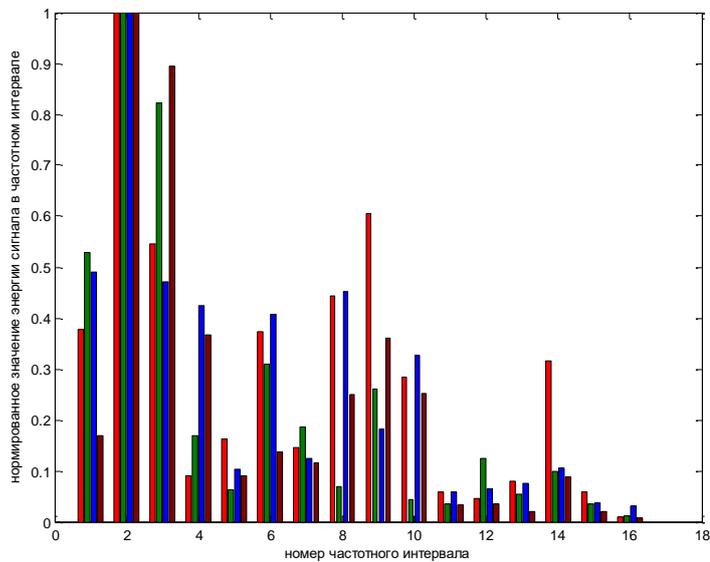
Б



В

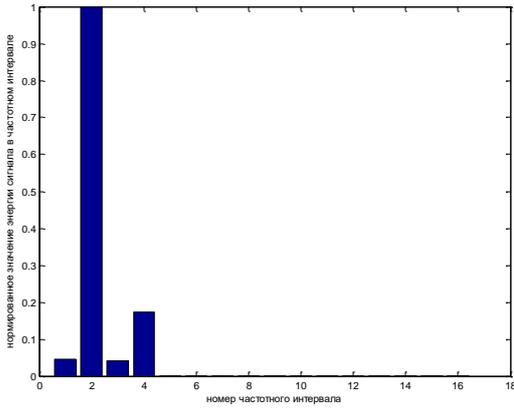


Г

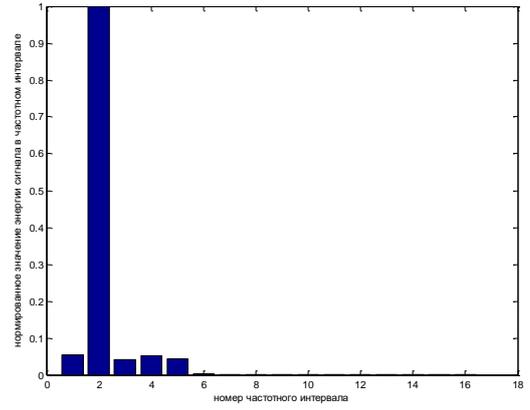


Д

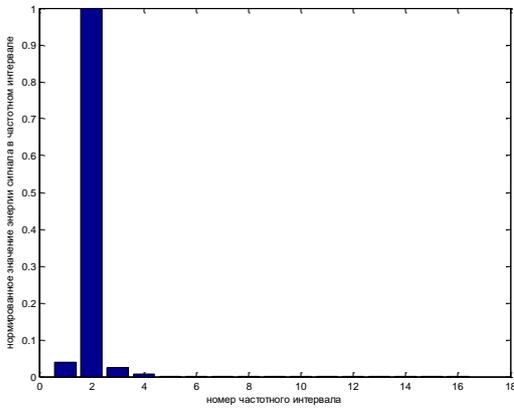
Рисунок 1.18 – Распределение энергии по частотным интервалам, где в качестве отрезка сигнала выступает фонема «т» из слова «четыре», по частотным интервалам Диктором 1 в четырёх различных случаях, и общий вид графиков (Д)



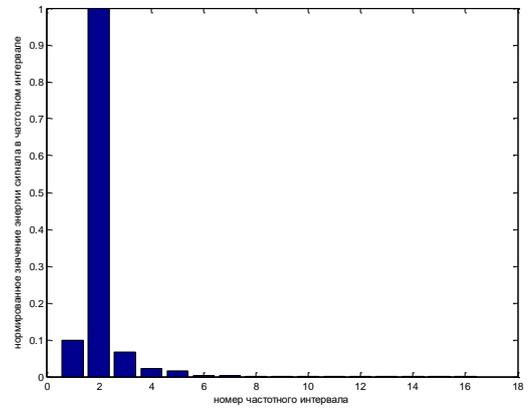
А



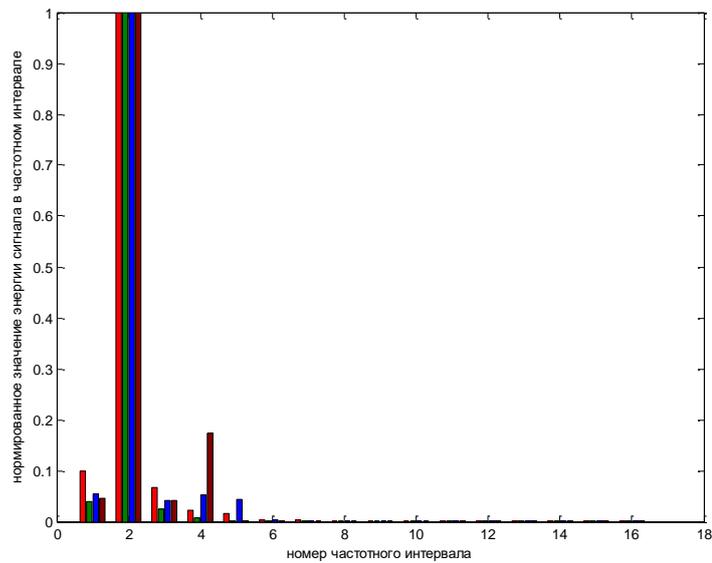
Б



В



Г



Д

**Рисунок 1.19 – Распределение энергии по частотным интервалам, где в качестве отрезка сигнала выступает фонема «о» из слова «восемь», по частотным интервалам Диктором 1 четырёх различных случаях, и общий вид графиков (Д)**

На основании рисунков 1.18 – 1.19 можно сделать вывод, что использование распределения энергии сигнала по частотным интервалам в качестве стабильных информативных признаков целесообразно, т.к. оно удовлетворяет условию стабильности на одних и тех же образцах голоса.

Кроме того, можно использовать информацию не о всей полосе частот, а только о тех частотных интервалах, в которых сосредоточена подавляющая доля энергии. Подобные частотные интервалы, удовлетворяющие (1.27), получили название информационных частотных интервалов [22, 25, 62].

$$\frac{\sum_{r \in R_1} P_r(\vec{x})}{\|\vec{x}\|^2} \approx 1, \quad (1.27)$$

где  $R_1$  – множество информационных частотных интервалов,  $R_1 \in R$ .

Для того чтобы отобрать значимые информационные интервалы, необходимо выбрать некоторый порог  $h_r$ , с которым сравнивается энергия в этом частотном интервале. В данной работе предлагается использовать в качестве порога значение "средней энергии" сигнала  $\vec{x}$ :

$$h_r = \frac{\Delta\omega}{\pi} \|\vec{x}\|^2 \quad (1.28)$$

Тогда  $\Omega_r$  - тый частотный интервал считается информационным, если он удовлетворяет условию (1.29):

$$P_r(\vec{x}) \geq h_r \quad (1.29)$$

### 1.3 Решающие правила

После получения стабильных информативных признаков диктора, полученный образ сравнивается с предварительно подготовленными эталонными образами, а затем принимается соответствующая логика принятия решений для определения голоса заданного диктора среди возможного множества [47]. Для этого вычисляется мера различимости для двух записей, на основе которой принимается и выносится решение.

В качестве мер близости наиболее часто используют метод вычисления расстояния  $d(\vec{x}, \vec{y})$ :

- евклидово расстояние:

$$d(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}, \quad i = 1, 2, \dots, N \quad (1.30)$$

где  $\vec{x}$  – вектор признаков, полученный из речевого сигнала пользователя, проходящего процедуру идентификации;

$\vec{y}$  – вектор признаков, хранимый в базе данных дикторов;

$x_i$  и  $y_i$  –  $i$ -тые элементы векторов  $\vec{x}$  и  $\vec{y}$ ;

$N$  – количество элементов векторов  $\vec{x}$  и  $\vec{y}$ .

Чем меньше значение  $d(\vec{x}, \vec{y})$ , тем более похожими являются вектора  $\vec{x}$  и  $\vec{y}$ .

- L1-норма (расстояние городских кварталов, манхэттэнское расстояние):

$$d(\vec{x}, \vec{y}) = \sum_{i=1}^N |x_i - y_i|, \quad i = 1, 2, \dots, N \quad (1.31)$$

где  $\vec{x}$  – вектор признаков, полученный из речевого сигнала пользователя, проходящего процедуру идентификации;

$\vec{y}$  – вектор признаков, хранимый в базе данных дикторов;

$x_i$  и  $y_i$  –  $i$ -тые элементы векторов  $\vec{x}$  и  $\vec{y}$ ;

$N$  – количество элементов векторов  $\vec{x}$  и  $\vec{y}$ .

Евклидово расстояние – наиболее общий тип расстояния [45].

Расстояние «городских кварталов» – по сравнению с евклидовым расстоянием влияние отдельных больших разностей (выбросов) уменьшается, так как они не возводятся в квадрат [45].

- метрика вычисления корреляции:

$$r = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|}, \quad (1.32)$$

где  $\vec{x}$  – центрированный вектор признаков, полученный из речевого сигнала пользователя, проходящего процедуру идентификации;

$\vec{y}$  – центрированный вектор признаков, хранимый в базе данных дикторов.

Чем больше значение  $r$ , тем более похожими являются вектора  $\vec{x}$  и  $\vec{y}$ .

Общим свойством всех норм является то, что обработке можно подвергать только векторы одной размерности. Это не всегда возможно, т.к. количество элементов вектора признаков зачастую зависит от длительности обрабатываемого речевого сигнала (трансформанта Фурье, кепстр), что

вносит необходимость исказить исходные данные (пр. – добавление в сигнал отсчётов с нулевой амплитудой для вычисления спектра) для получения необходимого количества признаков.

При решении задачи идентификации диктора происходит однократное сравнение индивидуальных признаков говорящего с соответствующими ему эталонами, на основе которого выносится положительное или отрицательное решение. В зависимости от вероятности ошибок I-го и II-го рода (т.е. верификация "самозванца" и отклонение "подлинного" диктора) устанавливается порог различимости. Решающее правило выглядит в соответствии (1.33):

$$\begin{aligned} & \text{Идентифицировать диктора } i, \text{ если } p_i(\vec{x}) \geq c_i p_{av}(\vec{x}) \\ & \text{Отклонить диктора } i, \text{ если } p_i(\vec{x}) < c_i p_{av}(\vec{x}), \end{aligned} \quad (1.33)$$

где  $p_i(\vec{x})$  – распределение вероятности индивидуальных признаков диктора;

$c_i$  – константа для  $i$  - того диктора, определяющая вероятность ошибок  $i$  - того диктора;

$p_{av}(\vec{x})$  – среднее (по всему ансамблю дикторов) распределение вероятности индивидуальных признаков диктора.

#### 1.4 Результаты и выводы главы. Задачи исследования

1. Сформулированы требования к выбору признаков для формирования признакового пространства и требования к выбору решающей функции, применяемые в задачах идентификации диктора;

2. Анализ различных подходов к формированию признакового пространства показал, что большинство вычисляемых признаков не

удовлетворяют условиям стабильности на одних и тех же образцах голоса, и резкому изменению при смене диктора. Использование распределения энергии сигнала по частотным интервалам в качестве стабильных информативных признаков является наиболее адекватным с точки зрения удовлетворения сформулированным требованиям.

3. Рассмотрены основные меры близости и сформулировано решающее правило, применимое в задачах идентификации дикторов. Отмечено, что важным критерием мер близости является то, что они могут быть вычислены только для векторов признаков одной длины.

На основании проведённого анализа состояния вопроса относительно методов идентификации дикторов и в соответствии с целью работы можно сформулировать следующие задачи исследования:

1. Анализ основных подходов, используемых в задачах идентификации дикторов;
2. Разработка решающей функции для идентификации диктора на основе субполосного подхода;
3. Разработка и исследование субполосного метода и алгоритма идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение частей энергии по частотным интервалам;
4. Разработка и исследование субполосного метода и алгоритма идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение долей энергии информационных частотных интервалов.
5. Проведение вычислительных экспериментов.

## ГЛАВА 2 РАЗРАБОТКА МЕТОДА СУБПОЛОСНОЙ ИДЕНТИФИКАЦИИ ДИКТОРОВ

В процессе исследования был разработан метод субполосной идентификации дикторов, который включает в себя решения следующих задач:

- 1) удаление пауз из речевого сигнала диктора;
- 2) сегментация речевого сигнала на однородные отрезки;
- 3) выбор решающей функции и построение критической области.

### 2.1 Удаление пауз из речевого сигнала диктора

В голосовой связи речь может быть охарактеризована как прерывистый носитель информации из-за пауз, которые являются характерными особенностями данного вида сигналов [10]. Интервалы, где голосовая информация присутствует, называют участками активной речи, а паузы между речевыми «всплесками» – участками пауз.

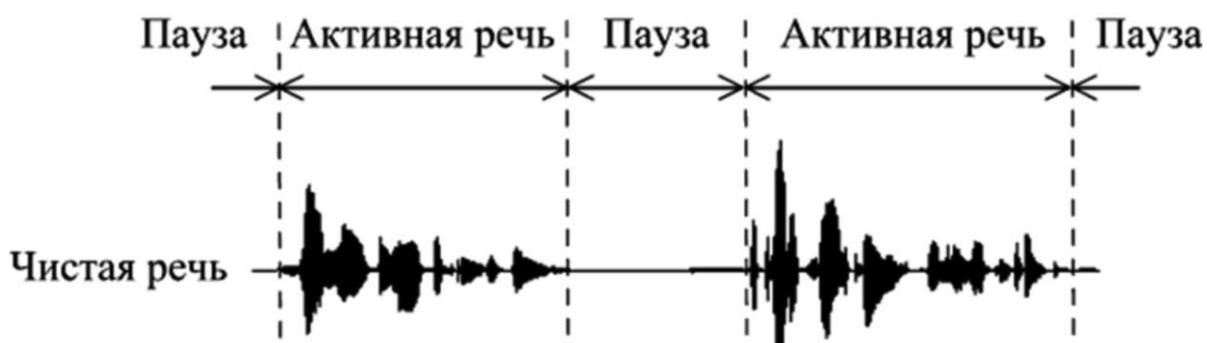


Рисунок 2.1 – Участки активной речи и паузы

Для того, чтобы производить анализ только на участках активной речи, воспользуемся описанным в работах [5, 23, 24] методом обнаружения пауз в

речевом сигнале, основанный на вариационном методе анализа речевых сигналов.

Распределение долей энергии по частотным интервалам может быть использовано для обнаружения пауз в речевом сигнале. Решающая функция для проверки гипотезы о том, что анализируемый отрезок сигнала соответствует паузе между звуками данных имеет вид [23, 24]:

$$S = \max(P_r / P_r^{\Pi}) \text{ для любых } r = 1, 2, \dots, R, \quad (2.1)$$

где  $P_r$  – значение части энергии  $r$ -ого частотного интервала для отрезка сигнала, соответствующего активной речи;

$P_r^{\Pi}$  – значение части энергии  $r$ -ого частотного интервала для отрезка сигнала, соответствующего паузе;

$S$  – решающая функция.

При этом значение  $P_r^{\Pi}$  получается на этапе обучения за счёт усреднения по большому количеству отрезков сигнала, заведомо относящимся к паузам:

$$P_r^{\Pi} = \frac{\sum_{k=1}^{N_y} (P_r)_k^{\Pi}}{N_y}, \quad (2.2)$$

где  $N_y$  – количество отрезков сигнала в паузе, которые используются для усреднения.

Если выполняется неравенство (2.3), то принимается, что анализируемый отрезок относится к паузе.

$$S > h_{\alpha}, \quad (2.3)$$

где  $h_\alpha$  – пороговое значение, обеспечивающее заданный уровень вероятности ложной тревоги, т.е.  $P\{S > h_\alpha\} \leq \alpha \ll 1$ .

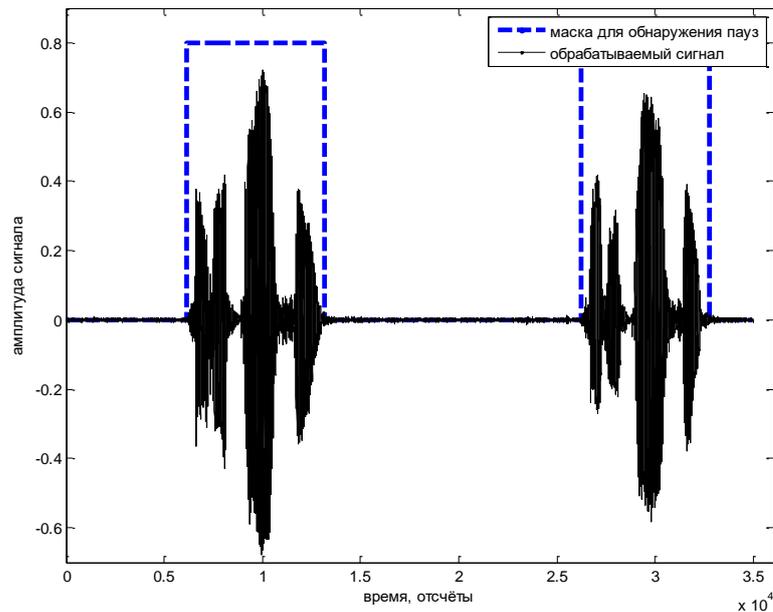
Пороговое значение  $h_\alpha$  можно получить из основанного на неравенстве Чебышева выражении (2.4).

$$h_\alpha \leq \bar{S}_\Pi + D_\Pi / d \cdot \sqrt{\alpha}, \quad (2.4)$$

где  $\bar{S}_\Pi$  – пороговое значение, обеспечивающее заданный уровень вероятности ложной тревоги, т.е.  $P\{S > h_\alpha\} \leq \alpha \ll 1$ .

Для проведения эксперимента был записан речевой материал, соответствующий фразе «черепаха черепаха». Речевой материал состоит не только из активной речи, но и из периода паузы, в самом начале записи (для обучающей выборки). Фраза была записана при частоте дискретизации 8000 Гц и 16 уровней квантования. Для удобства анализа была построена маска для обнаружения пауз, которая принимает значение «0» в случае, если анализируемый отсчёт речевого материала представляет собой шум. В противном случае – любое ненулевое постоянное значение (в данном эксперименте 0,8).

На рисунке 2.2 изображен результат работы реализованного алгоритма.



**Рисунок 2.2 – Результат работы алгоритма обнаружения пауз**

Таким образом на основании рисунка 2.2 можно сделать вывод, что метод обнаружения пауз в речевом сигнале, основанный на вариационном методе анализа речевых сигналов, может использоваться для поиска участков активной речи в задачах идентификации диктора.

После работы рассмотренного алгоритма все участки активной речи сохраняются как отдельные звуковые файлы для дальнейшего анализа.

## **2.2 Сегментация речевых сигналов на однородные отрезки**

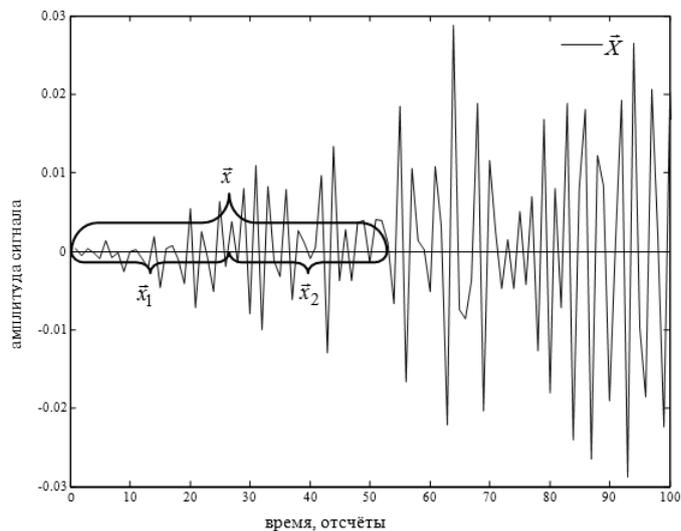
В определенном смысле базовым этапом автоматического анализа устной речи служит сегментация РС на отрезки, которые порождены в разных условиях [62]. Сегментация речевого сигнала – это процесс определения границ участков, соответствующих одному звуку или сочетанию звуков речи [22, 25, 62].

Сегментация на основе математического аппарата субполосного анализа и синтеза [22] учитывает свойство речевых сигналов, что распределения энергии меняются от звука к звуку.

Исходными данными является речевой сигнал («идентификатор»), регистрируемый на выходе микрофона под воздействием акустических колебаний, через АЦП ( $f_d = 8$  кГц, 16 уровней квантования), который записывается в память ПК в виде соответствующего звукового файла для последующего анализа.

Пусть  $\vec{X} = (X_1, \dots, X_N)$  – речевой сигнал, длительностью  $N$  отсчётов. Ось частот делится на  $(R+1)$  непересекающихся частотных интервала в соответствии с рисунком 2.3.

Из общего сигнала  $\vec{X}$  выделяются отрезки данных  $\vec{x}_1$  и  $\vec{x}_2$ , длиной  $NN = 2(2R+1)$  отсчётов (см. рисунок 2.3):



**Рисунок 2.3 – Выделение отрезков речевых данных во временной области**

Исходная (нулевая) гипотеза формируется следующим образом:

**$H_0$**  : два отрезка  $\vec{x}_1$  и  $\vec{x}_2$  порождены в одних и тех же условиях.

В качестве решающего правила предлагается использовать соотношение (2.5).

$$\frac{\sum_{r \in R_1} (\sqrt{P_r(\bar{x}_1)} - \sqrt{P_r(\bar{x}_2)})^2}{\|\bar{x}\|} \leq \frac{\left( \frac{B \cdot \Delta\omega}{\pi} \right) \|\bar{x}\|}{\sum_{r \in R_1} P_r(\bar{x})}, \quad (2.5)$$

где  $\bar{x}_1 = (X_k, \dots, X_{k+2NN-1})^T$ ,  $\bar{x}_2 = (X_{k+NN}, \dots, X_{k+2NN-1})^T$ ,  $k = 1, 2, \dots, (N - 2NN + 1)$ ,

$\bar{x} = [\bar{x}_1, \bar{x}_2]^T$  – отрезки речевых данных во временной области;

$P_r(\bar{x})$  – часть энергии, сосредоточенной в  $\Omega_r$  - ом частотном интервале сигнала  $\bar{x}$  (см. (1.24));

$R_1$  – множество информационных частотных интервалов (см. (1.27));

$B$  – количество информационных частотных интервалов;

$\Delta\omega$  – ширина  $\Omega_r$  - ого частотного интервала,  $r = 1, 2, \dots, R$  (см. (1.20));

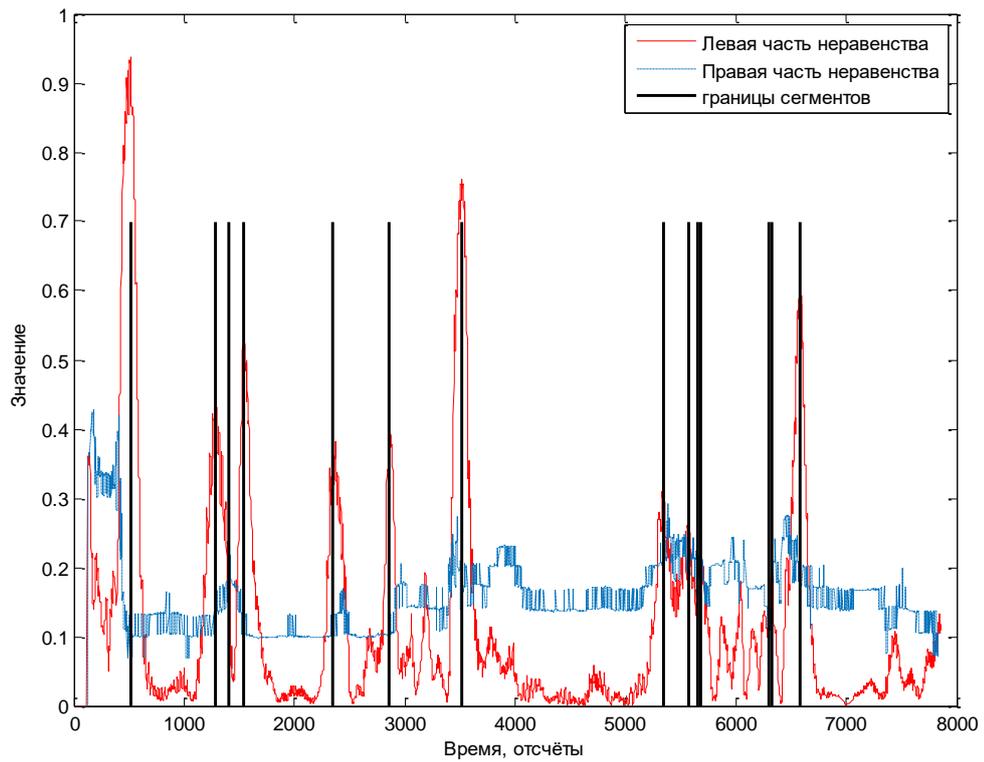
$\|\bar{x}\|$  – энергия отрезка речевого сигнала.

Левая часть неравенства (2.5) представляет собой решающую функцию, которая определяет расстояние между двумя отрезками на основе вычисления евклидова расстояния между корнями частей энергии значимых частотных интервалов.

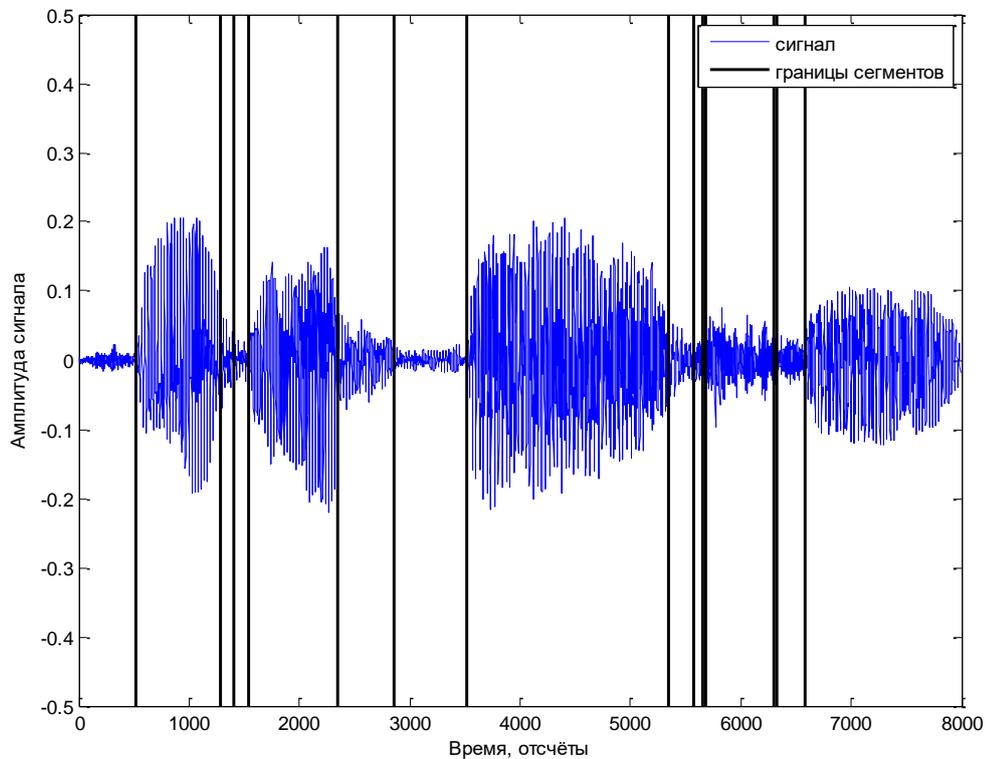
Правая часть неравенства (2.5) – порог, который при выполнении нулевой гипотезы, должен ограничивать значение решающей функции.

Левая и правая части неравенства вычисляются по - отдельности для каждой пары отрезков  $\bar{x}_1$  и  $\bar{x}_2$ , выбираемых по методу скользящего окна.

В случае, когда неравенство (2.5) не выполняется, то гипотеза  $H_0$  отвергается и ставится метка, по максимальному значению решающей функции. Эта метка и является границей между сегментами.



**Рисунок 2.4 – Решающее правило для фразы «черепаха»**



**Рисунок 2.5 – Сегментация речевого сигнала, соответствующего фразе «черепаха»**

Результатом работы алгоритма является вектор  $\vec{S} = (s_1, \dots, s_j, \dots, s_M)$ ,

где  $s_j$  - номер отсчёта, соответствующий границе сегмента;

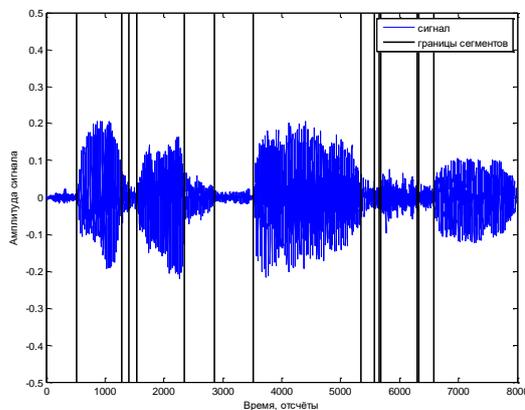
$s_1 = 1$  - номер отсчёта, соответствующий началу первого сегмента;

$s_M = N$  - номер отсчёта, соответствующий концу последнего сегмента;

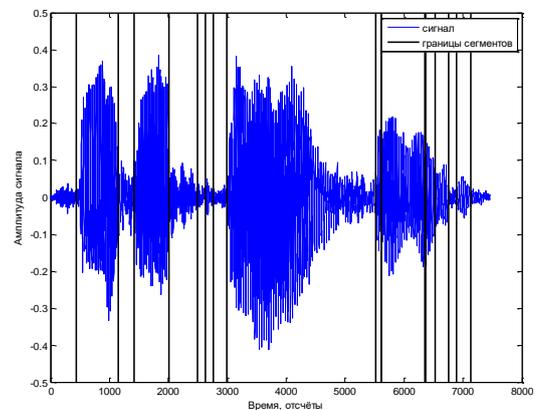
$(s_2, \dots, s_{M-1})$  - номера отсчётов, соответствующие границам сегментов,

выделенные алгоритмом (рисунки 2.4 – 2.5).

Особенностью предложенного метода сегментации является то, что при произнесении одной и той же фразы одним и тем же диктором, количество сегментов и их границы могут отличаться. Это может быть связано с тем, что, произнося одинаковую фразу, у диктора может меняться темп речи, громкость произношения, или его эмоциональное состояние.

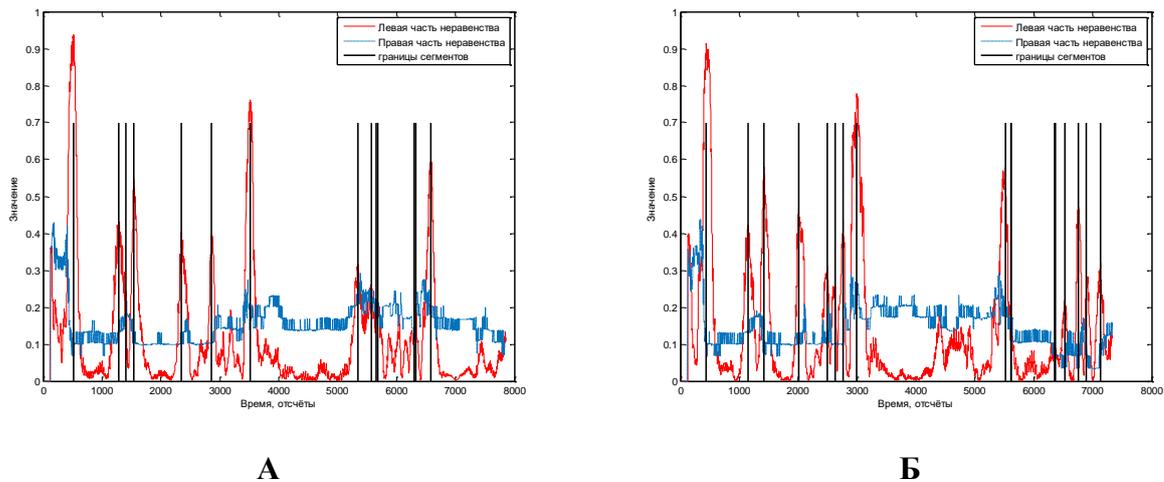


А



Б

**Рисунок 2.6 – Сегментация разных реализаций речевых сигналов, соответствующих фразе «черепаша», произнесённых одним диктором**



**Рисунок 2.7 – Вид решающей функции для фраз «один», произнесённых одним диктором**

### 2.3 Решающая функция для субполосной идентификации дикторов

Предлагается, что идентификация осуществляется посегментно.

Пусть имеются два сигнала:

$\vec{X} = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_i, \dots, \vec{x}_L]$  – речевой сигнал диктора, проходящего процедуру идентификации, где  $\vec{x}_i$  -  $i$ -тый сегмент сигнала  $\vec{X}$  ;

$\vec{Y} = [\vec{y}_1, \vec{y}_2, \dots, \vec{y}_k, \dots, \vec{y}_M]$  – речевой сигнал диктора, с которым происходит сравнение, где  $\vec{y}_k$  -  $k$ -тый сегмент сигнала  $\vec{Y}$  ;

Основная гипотеза формулируется следующим образом:

**H<sub>0</sub>**: сигналы  $\vec{X}$  и  $\vec{Y}$  соответствуют разным дикторам.

В данной работе будут рассмотрены решающие функции для двух случаев:

1) когда в качестве стабильных информативных признаков используется распределение энергии сигнала по частотным интервалам;

2) когда в качестве стабильных информативных признаков используется распределение долей энергии информационных частотных интервалов.

### 1) Использование в качестве признаков распределение энергии сигнала по частотным интервалам

В качестве решающей функции предлагается использовать коэффициент соответствия  $F^z(\vec{X}, \vec{Y}^z)$  (2.6):

$$F^z(\vec{X}, \vec{Y}^z) = \left( \frac{\sum_{i=1}^L \sum_{r=1}^R (P_r(\bar{x}_i) \cdot P_r(\bar{y}_k^z))}{\|\bar{x}_i\| \cdot \|\bar{y}_k^z\|} \right), z = 1, 2, \dots, Z, \quad (2.6)$$

где  $Z$  – количество речевых сигналов, зарегистрированных в системе идентификации.

Нулевая гипотеза  $H_0$  отвергается при выполнении условия (2.7):

$$F^z(\vec{X}, \vec{Y}^z) > h \quad (2.7)$$

Для установления порога  $h$  осуществляется оценка критической области при заданной вероятности ошибки правильной идентификации  $P_{ош}$ :

$$PR(F^z(\vec{X}, \vec{Y}^z) > h / H_0) \leq P_{ош} \quad (2.8)$$

### 2) Использование в качестве признаков распределение долей энергии информационных частотных интервалов

В качестве решающей функции предлагается использовать выражение (2.9):

$$Fd^z(\vec{X}, \vec{Y}^z) = \left( \frac{\sum_{r=1}^R Pd_r(\vec{x}_i) \cdot Pd_r(\vec{y}_k^z)}{\left( \sum_{r=1}^R Pd_r(\vec{x}_i) \right) \cdot \left( \sum_{r=1}^R Pd_r(\vec{y}_k^z) \right)} \right), \quad z = 1, 2, \dots, Z \quad (2.9)$$

где  $Z$  – количество речевых сигналов, зарегистрированных в системе идентификации;

$Pd_r(\vec{x})$  – значение доли энергии информационного частотного интервала (см. (1.27)):

$$Pd_r(\vec{x}) = \begin{cases} P_r(\vec{z}) / \|\vec{x}\|, & \text{если } r \in R_1 \\ 0 & \text{если } r \notin R_1 \end{cases} \quad (2.10)$$

Нулевая гипотеза  $H_0$  отвергается при выполнении условия (2.10):

$$Fd^z(\vec{X}, \vec{Y}^z) > h \quad (2.11)$$

Для установления порога  $h$  осуществляется оценка критической области при заданной вероятности ошибки правильной идентификации  $P_{oui}$ :

$$PR(Fd^z(\vec{X}, \vec{Y}^z) > h / H_0) \leq P_{oui} \quad (2.12)$$

## 2.4 Результаты и выводы главы

1. Реализован адаптивный алгоритм удаления пауз, которая может быть применена для селекции потока речи на отдельные слова;

2. Реализован адаптивный алгоритм сегментации речевого сигнала на однородные отрезки, что может быть использовано для поиска характеристик, неизменных в пределах анализируемого сегмента;

3. Разработана решающая функция для субполосный метода идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение частей энергии по частотным интервалам;

4. Разработана решающая функция для субполосного метода идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение долей энергии информационных частотных интервалов.

## ГЛАВА 3 ИССЛЕДОВАНИЕ МЕТОДА СУБПОЛОСНОЙ ИДЕНТИФИКАЦИИ ДИКТОРОВ

### 3.1 Реализация этапа обучения

Этап обучения проводится для того, чтобы определить пороговое значение, относительно которого принимается решение об идентификации.

Идея алгоритма заключается в том, что значение коэффициента сопоставления для записей, соответствующий одному и тому же диктору  $F^z(\vec{X}, \vec{X}^z)$ , будет больше, чем значение коэффициента сопоставления для записей двух разных дикторов  $F^z(\vec{X}, \vec{Y}^z)$ . Поэтому можно предположить, что существует некоторое пороговое значение, которое  $F^z(\vec{X}, \vec{Y}^z)$  не может превысить.

Если рассматривать случай, когда вероятность ошибки II рода должна быть равна нулю, то имеет смысл принимать за порог  $h$  максимальное значение  $F^z(\vec{X}, \vec{Y}^z)$ . Т.е. возможны два варианта:

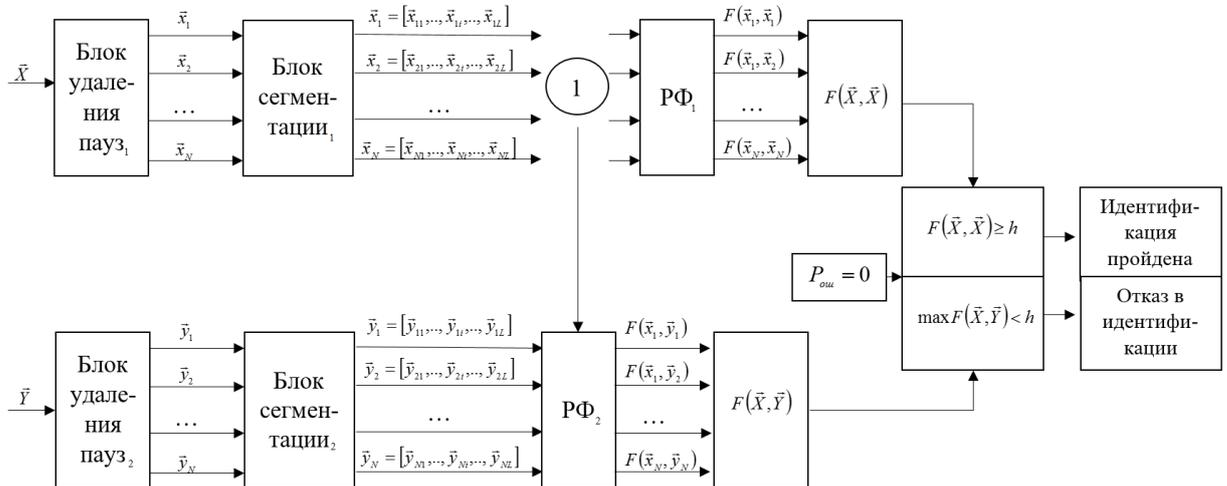
$$\begin{aligned} F^z(\vec{X}, \vec{X}^z) \geq h, & \quad \text{идентификация пройдена} \\ \max F^z(\vec{X}, \vec{Y}^z) < h, & \quad \text{отказ в идентификации} \end{aligned} \quad (3.1)$$

Смыслом (3.1) является то, что идентификация считается пройденной успешно только в случае, если значение  $F^z(\vec{X}, \vec{X}^z) \in [h, 1]$ .

Для удобства использования рассмотренных в главе 2 алгоритмов была разработана функциональная схема метода субполосной идентификации дикторов (рисунок 3.1), которая включает в себя

- 1) блок удаления пауз;
- 2) блок сегментации речевых сигналов на однородные отрезки;

- 3) блок вычисления решающей функции;
- 4) блок принятия решения.



**Рисунок 3.1 – Функциональная схема метода субполосной идентификации дикторов**

$\vec{X}$  – речевой сигнал диктора, проходящего процедуру идентификации для построения обучающей выборки, содержащий многократные реализации произнесения парольного слова. РС  $\vec{X}$  поступает на вход блока удаления пауз<sub>1</sub> и селекции потока речи на отдельные слова. Выходными данными блока являются набор речевых сигналов  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$ , где  $N$  – количество произнесения диктором парольного слова.

В блоке сегментации<sub>1</sub> определяются границы участков сегментов с использованием РФ, основанной на учете субполосного расстояния. На выходе данного блока речевые сигналы  $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$  представляются в виде наборов сегментов:  $\vec{x}_1 = [\vec{x}_{11}, \dots, \vec{x}_{1l}, \dots, \vec{x}_{1L}]$ ,  $\vec{x}_2 = [\vec{x}_{21}, \dots, \vec{x}_{2l}, \dots, \vec{x}_{2L}]$ , ...,  $\vec{x}_N = [\vec{x}_{N1}, \dots, \vec{x}_{Ni}, \dots, \vec{x}_{NL}]$ .

Одновременно аналогичные операции происходят и с речевым сигналом другого диктора  $\vec{Y}$ .

На вход блока  $R\Phi_1$  поступают наборы сегментов:  $\vec{x}_1 = [\vec{x}_{11}, \dots, \vec{x}_{1i}, \dots, \vec{x}_{1L}]$ ,  $\vec{x}_2 = [\vec{x}_{21}, \dots, \vec{x}_{2i}, \dots, \vec{x}_{2L}]$ , ...,  $\vec{x}_N = [\vec{x}_{N1}, \dots, \vec{x}_{Ni}, \dots, \vec{x}_{NL}]$ , для которых вычисляются значения коэффициентов сопоставления  $F(\vec{x}_1, \vec{x}_1)$ ,  $F(\vec{x}_1, \vec{x}_2)$ , ...,  $F(\vec{x}_N, \vec{x}_N)$ , которые удобно сохранять в виде матрицы  $F^z(\vec{X}, \vec{X}^z)$ , где  $Z = 1, 2, \dots, N$ .

На вход блока  $R\Phi_2$  поступают наборы сегментов для речевых сигналов двух разных дикторов. Результатом работы данного сегмента являются значения коэффициентов сопоставления  $F(\vec{x}_1, \vec{y}_1)$ ,  $F(\vec{x}_1, \vec{y}_2)$ , ...,  $F(\vec{x}_N, \vec{y}_N)$ , которые удобно сохранять в виде матрицы  $F^z(\vec{X}, \vec{Y}^z)$ , где  $Z = 1, 2, \dots, N$ .

В случае, рассмотренном на рисунке 3.1, а именно, когда вероятность ошибок II рода равна 0, пороговое значение вычисляется как:

$$\max F^z(\vec{X}, \vec{Y}^z) + c = h, \quad (3.2)$$

где  $c$  – константа,  $c > 0$ .

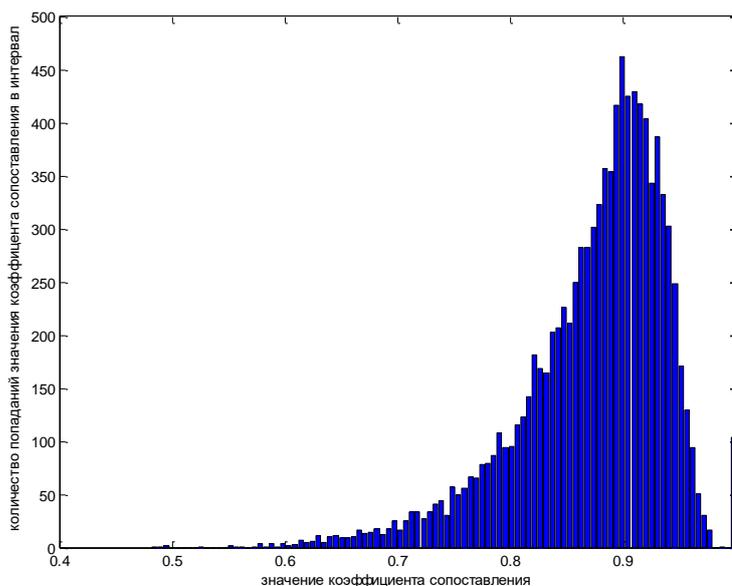
После чего принимается решение об идентификации в соответствии с (3.1).

### **3.2 Вычисление порогового значения для метода субполосной идентификация дикторов, где качестве признаков используется распределение энергии сигнала по частотным интервалам**

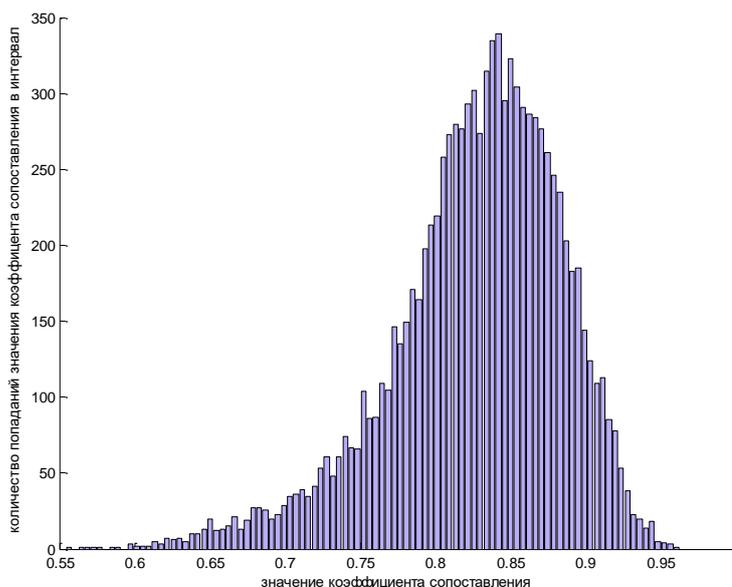
Для того, чтобы наглядно продемонстрировать работу метода субполосной идентификации дикторов, построим гистограмму распределений коэффициентов соответствий  $F^z(\vec{X}, \vec{X}^z)$  (обучающая выборка) (рисунок 3.2) и гистограмму  $F^z(\vec{X}, \vec{Y}^z)$  (выборка идентификации) (рисунок 3.3), где при вычислении решающей функции используется вектор

распределения энергии сигнала по частотным интервалам в качестве стабильных информативных признаков.

Для реализации эксперимента была создана база из пяти дикторов, каждому из которых соответствует по 100 реализаций произнесения слова «черепаха» (частота дискретизации 8000 Гц, 16 бит квантования)

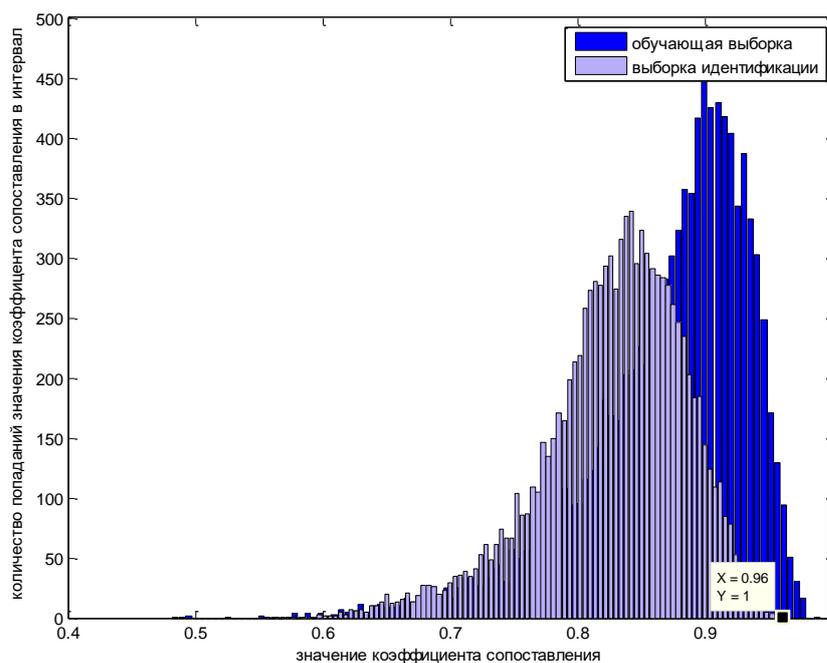


**Рисунок 3.2** Гистограмма распределения значений коэффициента сопоставления  $F^z(\vec{X}, \vec{X}^z)$ , где в качестве  $\vec{X}$  выступает Диктор 1



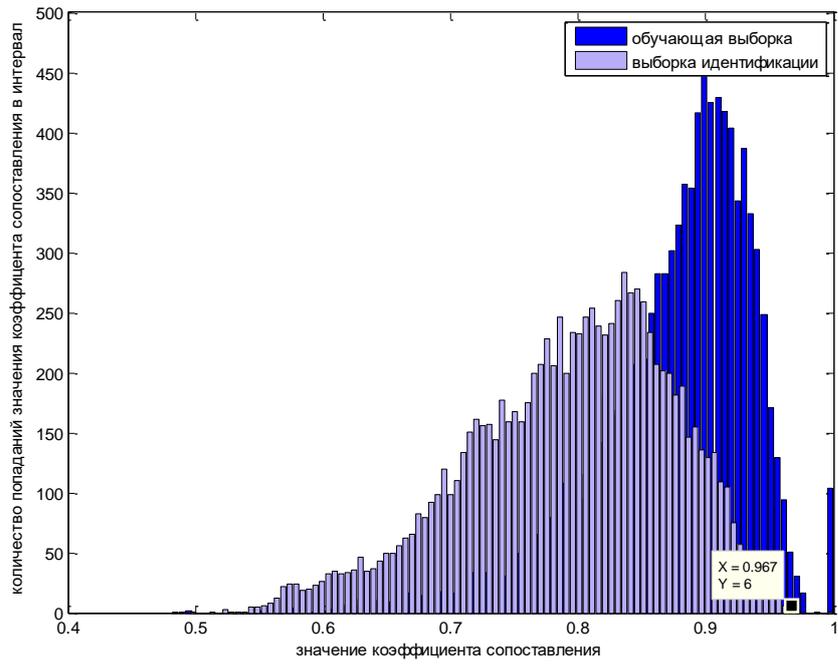
**Рисунок 3.3** Гистограмма распределения значений коэффициента сопоставления  $F^z(\vec{X}, \vec{Y}^z)$ , где в качестве  $\vec{X}$  выступает Диктор 1,  $\vec{Y}$  - Диктор 2

Если совместить данные с рисунков 3.2 и 3.3, можно найти крайнее значение порога, относительно которого происходит решение об идентификации «свой/чужой». Под крайним значением порога будем понимать такое значение, при незначительном увеличении которого на константу  $c$  может быть обеспечена нулевая вероятность ошибок II рода.

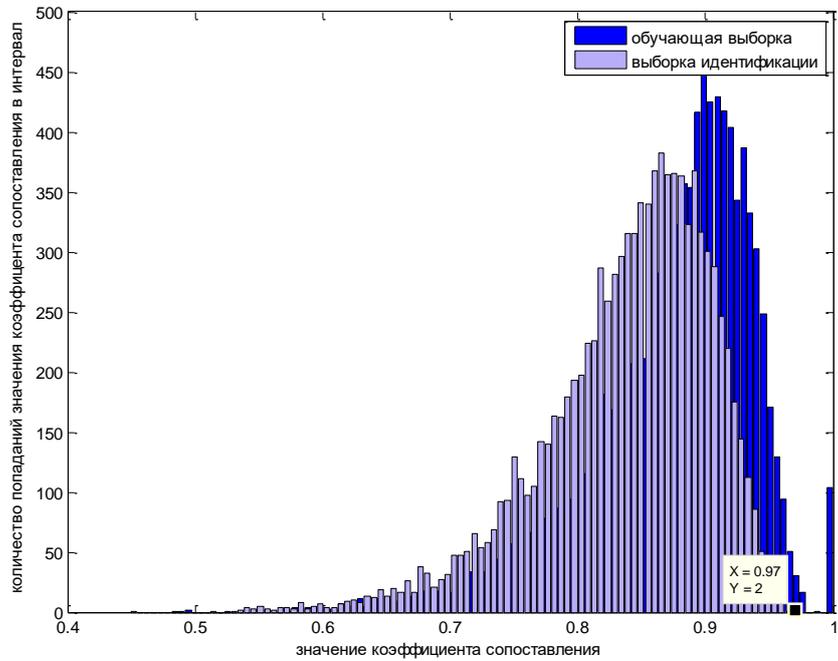


**Рисунок 3.4 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 1**

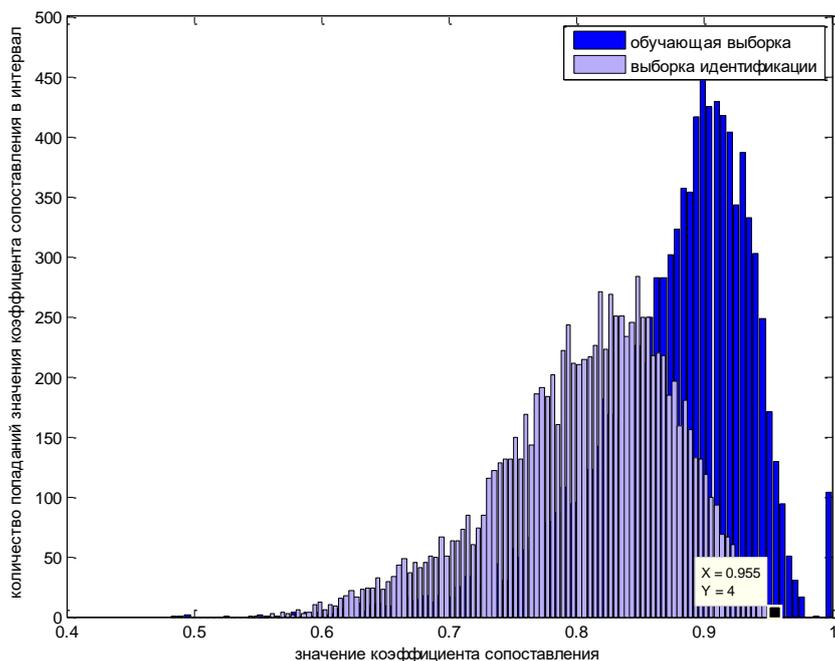
На рисунках 3.5 – 3.7 представлены аналогичные гистограммы, только при прохождении процедуры идентификации другими дикторами 3-5.



**Рисунок 3.5 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 1**



**Рисунок 3.6 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 1**



**Рисунок 3.7 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 1**

Если повторить подобные эксперименты для каждого из дикторов, то все данные можно свести в таблицу 3.1, где указаны крайние пороговые значения для каждой пары дикторов. Гистограммы для этих случаев представлены в ПРИЛОЖЕНИИ А.

**Таблица 3.1 – Пороговые значения для метода субполосной идентификация дикторов, где качестве признаков используется распределение энергии сигнала по частотным интервалам**

		Выборка идентификации				
		Диктор 1	Диктор 2	Диктор 3	Диктор 4	Диктор 5
1		2	3	4	5	6
Обучающая выборка	Диктор 1		0.96	0.967	0.97	0.955
	Диктор 2	0.932		0.952	0.954	0.946
	Диктор 3	0.965	0.964		0.971	0.974

Продолжение таблицы 3.1

	1	2	3	4	5	6
Обучающая выборка	Диктор 1	0.955	0.958	0.958		0.959
	Диктор 5	0.948	0.961	0.964	0.965	

Из таблицы 3.1 можно сделать вывод, что крайние пороговые значения не превышают величину 0,974, которая и может использоваться для задания порогового значения критической области принятия решений, полученное при обучении.

### 3.3 Вычисление порогового значения для метода субполосной идентификация дикторов, где качестве признаков используется распределение долей энергии информационных частотных интервалов

Если повторить исследования, аналогичные указанным в п.3.2, (см. ПРИЛОЖЕНИЕ Б), то получим таблицу 3.2.

Таблица 3.2 – Пороговые значения для метода субполосной идентификация дикторов, где качестве признаков используется распределение долей энергии информационных частотных интервалов

		Выборка идентификации				
		Диктор 1	Диктор 2	Диктор 3	Диктор 4	Диктор 5
Обучающая выборка	Диктор 1		0.95	0.961	0.964	0.95
	Диктор 2	0.927		0.949	0.947	0.941
	Диктор 3	0.959	0.96		0.966	0.971
	Диктор 4	0.955	0.95	0.956		0.949
	Диктор 5	0.946	0.961	0.963	0.962	

При сравнении значений из таблиц 3.1 и 3.2 можно отметить, что крайние пороговые значения для метода субполосной идентификации дикторов, где в качестве признаков используется распределение долей энергии информационных частотных интервалов, меньше, чем для в случае использования в качестве признаков распределения энергии отрезка по частотным интервалам. Подобный подход позволяет говорить о большей эффективности использования подобных признаков для решения задач идентификации дикторов.

Максимальное крайнее пороговое значение может приниматься за порог, на основе которого и будет выноситься решение об успешной/неуспешной идентификации.

Проведенные сравнительные эксперименты свидетельствуют о том, что предлагаемые стабильные информативные признаки и разработанные решающие функции могут найти своё применение в задачах идентификации диктора.

### **3.4 Результаты и выводы главы**

1. Исследовано поведение решающей функции и вычислены пороговые значения для ограничения критической области для субполосного метода идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение частей энергии по частотным интервалам;

2. Исследовано поведение решающей функции и вычислены пороговые значения для ограничения критической области для субполосного метода идентификации дикторов, где в качестве устойчивых информативных признаков используются распределение долей энергии информационных частотных интервалов.

3. Установлено, что пороговое значения разработанной решающей функции, когда в качестве стабильных информативных признаков используется распределение долей энергии информационных частотных интервалов, меньше, чем пороговое значение решающей функции, когда в качестве стабильных информативных признаков используется распределение энергии сигнала по частотным интервалам. Подобный вывод позволяет говорить о большей эффективности использования подобных признаков для решения задач идентификации дикторов.

## ЗАКЛЮЧЕНИЕ

1. Сформулированы требования к признакам, на основе которых формируется признаковое пространство для идентификации дикторов, и используемым в этих задачах решающим функциям. Проведён сравнительный анализ, насколько признаки, получаемые на основе спектрального, кепстрального и субполосного подходов, удовлетворяют сформулированным требованиям и адекватно отражают свойства обрабатываемых речевых сигналов. Отмечено, что использование распределения энергии сигнала по частотным интервалам (субполосный подход) в качестве стабильных информативных признаков является наиболее адекватным с точки зрения удовлетворения этим требованиям. Также предложена идея использовать в качестве стабильных информативных признаков распределение долей энергии информационных частотных интервалов, для повышения стабильности работы алгоритмов идентификации диктора.

2. Изложены и реализованы основные этапы, входящие в метод субполосной идентификации дикторов. В основе каждого из этапов лежит субполосный подход, который позволяет реализовывать адаптивные алгоритмы обработки речевых данных. Разработаны решающие функции для метода субполосной идентификации дикторов, где в качестве признаков, характеризующих особенности голоса диктора, могут выступать как распределение энергии сигнала по частотным интервалам, так и распределение долей энергии информационных частотных интервалов. Кроме того, задаётся понятие порога для ограничения критической области, на основе которого будет осуществляться решение об идентификации при заданной вероятности ошибки II рода.

3. Исследованная решающая функция построена таким образом, чтобы вероятность попадания значения решающей функции в критическую область была достаточно мала. Это было обосновано путём проведения ряда предварительных экспериментов, формируя обучающую выборку. Целью этих экспериментов было установление порогового значения решающей функции, относительно которого принимается решение об идентификации (отвергается первоначальная гипотеза  $H_0$ ). При этом значение порога может варьироваться в зависимости от необходимой эффективности работы алгоритма, а именно, насколько мала должна быть вероятность ошибки II рода.

4. Было установлено, что пороговое значения разработанной решающей функции, когда в качестве стабильных информативных признаков используется распределение долей энергии информационных частотных интервалов, меньше, чем пороговое значение решающей функции, когда в качестве стабильных информативных признаков используется распределение энергии сигнала по частотным интервалам. Подобный вывод позволяет говорить о большей эффективности подобных признаков для решения задач идентификации дикторов.

5. Проведенные сравнительные эксперименты свидетельствуют о том, что предлагаемые стабильные информативные признаки и разработанные решающие функции могут найти своё применение в задачах идентификации диктора. Поэтому из полученных результатов исследования следует утверждать, что разработанные методы и алгоритмы позволяют усовершенствовать существующие методы и алгоритмов идентификации дикторов по голосу.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Агашин, О.С. Методы цифровой обработки речевого сигнала в задаче распознавания изолированных слов с применением сигнальных процессоров / О.С.Агашин, О.Н.Корелин // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева – 2012. - №4 (97). – С.32 –44.
2. Аграновский, А.В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов [Текст]/ А.В. Аграновский, Д.А. Леднов – М.: Радио и связь, 2004. – 164 с.
3. Айфичер Э. Цифровая обработка сигналов: практический подход, 2-е издание [Текст]: Пер. с англ. / Айфичер Э., Джервис Б. – М.: Издательский дом «Вильямс», 2004. – 992 с.:ил.
4. Барнет Дж. А. Методы автоматического распознавания речи: В 2-х книгах. [Текст]: Пер. с англ. / Дж.А. Барнет, М.И. Бернштейн и др. Под ред. У. Ли. – М.: Мир, 1983. – Кн. 2. 392 с., ил.
5. Белов С.П. Об уменьшении объема трафика при пакетной передаче речевых сообщений за счет кодирования пауз [Текст] / С.П. Белов, Е.И. Прохоренко // Научные ведомости Белгородского государственного университета. Серия: Информатика и прикладная математика. - Белгород, 2006. - № 1 (21). – Вып. 2. - С. 141–148.
6. Богданович, В.А. Теория устойчивого обнаружения, различения и оценивания сигналов. – 2-е изд., испр. [Текст]/ В.А. Богданович, А.Г. Вострецов.-М.: ФИЗМАТЛИТ, 2004. – 320 с.
7. Бондарев, В.Н. Цифровая обработка сигналов. Методы и средства [Текст]: Учеб. Пособие для вузов. 2-е изд / Бондарев В.Н., Трестре Г., Чернега В.С. – Х: Конус, 2001. – 398 с.:ил.

8. Вапник, В.Н. Теория распознавания образов (статистические проблемы обучения) [Текст]/ В.Н. Вапник, А.Я. Червоненкис.- Издательство «Наука», 1974. – 416 с.

9. Винцюк, Т.К., Анализ, распознавание и интерпретация речевых сигналов [Текст]/ Т.К. Винцюк - Киев: Наук.думка, 1987. - 264с.

10. Волченков В. А. Методы и алгоритмы детектирования активности речи [Текст] / В.А. Волченков, В.В. Витязев // Цифровая Обработка Сигналов. – 2013. - № 1. – С. 54-60.

11. Глинченко, А. С. Цифровая обработка сигналов [Текст]: Учеб. пособие: В 2 ч. Ч. 2. / А.С. Глинченко. – Красноярск: ИПЦ КГТУ, 2001. 184 с.

12. Голд, Б. Цифровая обработка сигналов [Текст] / Б. Голд, Ч. Рейдер. – М.: Сов. Радио, 1973. – 368 с.

13. Гольденберг Л.М. Цифровая обработка сигналов. [Текст] / Л.М. Гольденберг. – М.: Радио и связь, 1985. – 312 с.

14. Гоноровский И.С. Радиотехнические цепи и сигналы. Учебник для вузов. – 4-е изд., перераб. и доп. [Текст] / И.С. Гоноровский.– М.: Радио и связь, 1977. – 512 с.:ил.

15. Грибанов Ю.И. Спектральный анализ случайных процессов [Текст] / Ю.И. Грибанов, В.Л. Мальков. – М.: «Энергия», 1974. – 240 с.:ил.

16. Гудонавичюс, Р.В. Распознавание речевых сигналов по их структурным свойствам [Текст]/Р.В. Гудонавичюс, П.П. Кемешис, А.Б. Читавичюс – Л.: «Энергия», 1977. – 64 с.

17. Динамические спектры речевых сигналов [Текст] М.Ф. Деркач, Р.Я. Гумецкий, Б.М. Гура, М.Е. Чабан. – Львов: Вища школа. Изд-во при Львов. Ун-те, 1983. - 168 с.

18. Заковряшин, А.С. Применение распределений мел-частотных кепстральных коэффициентов для голосовой идентификации личности [Текст] / А.С. Заковряшин, П. В. Малинин, А. А. Лепендин // Известия

Алтайского Государственного университета, Том 1 – 2014. - №1 (81). – С.156 –160.

19. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. [Текст] / Е.Г. Жилияков. - Белгород: Изд-во БелГУ, 2007. – 160 с.

20. Жилияков Е.Г. Методы анализа и построения функций по эмпирическим данным на основе частотных представлений: монография [Текст] / Е.Г. Жилияков. - Белгород: Изд-во БелГУ, 2007. – 160 с.

21. Жилияков Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений [Текст] / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко. – Белгород: БелГУ, 2007. – 136 с.

22. Жилияков Е. Г. О сегментации речевых сигналов на однородные отрезки [Текст] / Е.Г. Жилияков, С.П. Белов, А.С. Белов, А.А. Фирсова // Научные ведомости белгородского государственного университета. Серия: Экономика. Информатика . Том 34 – 2015. - №7-1. – С. 194-199.

23. Жилияков Е. Г. Об анализе и синтезе речевых сигналов в IP-телефонии [Текст] / Е.Г. Жилияков, А.А. Фирсова // Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование. – 2009. - № 43. – С. 84-91.

24. Жилияков Е.Г. Об эффективности различных подходов, к сегментации речевых сигналов на основе обнаружения пауз [Текст] / Е.Г. Жилияков, С.П. Белов, А.С. Белов и др. ; БелГУ // Научные ведомости БелГУ. Сер. История. Политология. Экономика. Информатика. - 2010. - №7(78), вып.14/1.-С. 187-193.

25. Жилияков, Е.Г. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Научные ведомости Белгородского государственного университета. Серия: История.

Политология. Экономика. Информатика, Том 18 – 2011. - №7-1 (102). – С. 187-196.

26. Залманзон Л.А. Преобразование Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. [Текст] / Залманзон Л.А. – М.: Наука. Гл. ред. Физ.-мат.лит., 1989 год. 496 с.

27. Карпов, Н.В. Автоматический анализ качества устной речи [Текст] / Н.В. Карпов // Вестник Нижегородского университета им. Н.И.Лобачевского. – 2013. - №1(1). – С.219-224.

28. Кисиленко А.В. Субполосная идентификация диктора [Текст]/ А.В. Кисиленко // XIII-я Курчатовская молодёжная научная школа: сборник аннотаций 27 – 30 октября 2015 г. – Москва, 2015. – С. 158.

29. Кисиленко А.В. Субполосное распознавание дикторов [Текст] / А.В. Кисиленко, Д.И. Трубицына // Проблемы передачи и обработки информации в сетях и системах телекоммуникаций: Материалы 18-й Международной науч.-техн. конф.: Горячая-линия Телеком, 2015. – С.76–78.

30. Кипяткова, И.С. Автоматическая обработка разговорной русской речи: монография / И.С. Кипяткова, А.Л. Ронжин, А.А. Карпов. СПИИРАН. – СПб.: ГУАП, 2013. – 314 с.

31. Компьютерные технологии распознавания речевых сигналов [Текст] / О.Н. Карпов, А.Г. Габович, Б.Г. Марченко и др. – К.: ООО «ПолиграфКонсалтинг», 2005. – 138 с.

32. Котомин, А.В. Распознавание речевых команд с использованием сверточных нейронных сетей/ А.В. Котомин // Научно-информационные технологии // Труды XVI Молодежной научно-практической конференции SIT-2012 : г. Переславль-Залесский : апрель 2011 : УГП имени А. К. Айламазяна. – 2012. – С.17-28.

33. Кудашев, О.Ю. Система разделения дикторов на основе вероятностного линейного дискриминантного анализа. Диссертация на

соискание учёной степени кандидата технических наук — Санкт-Петербург, 2014. — 158 с.

34. Куприянов, М.С. Цифровая обработка сигналов: процессоры, алгоритмы, средства проектирования [Текст] / М.С. Куприянов. – СПб.: Политехника, 1999. – 592 с.

35. Лайонс, Р. Цифровая обработка сигналов / Лайонс Р; - 2-е изд. ; Пер. с англ. – М.: ООО "Бином-Пресс", 2006 – 656 с.: ил.

36. Марпл-мл, С.Л. Цифровой спектральный анализ и его приложения / Марпл-мл. С.Л.; Пер. с англ. –М.: Мир, 1990.

37. Методы цифровой обработки сигналов для решения прикладных задач. Монография [Текст] / Под ред. В.И. Марчука. – М.: Радиотехника, 2012. – 128 с.:ил.

38. Михайлов В.Г., Златоусов Л.В. Измерение параметров речи [текст]/ В.Г. Михайлов, Л.В. Златоусова; Под.ред. М.А. Сапожникова. – Москва: Радио и связь, 1987. – 168с.: ил.

39. Мясникова Е.Н. Объективное распознавание звуков речи [Текст] / Е.Н. Мясникова. – Л.: Энергия, 1967. – 150 с.: рис.

40. Назаров М.В. Методы цифровой обработки и передачи речевых сигналов [Текст] / М.В. Назаров, Ю.Н. Прохоров. – М.: Радио и связь, 1985. – 176 с.

41. Николенко, С. Лекция 6. Признаки. Кепстральные коэффициенты. MFCC / Е.Селифонов, А.Тихомиров / [Электронный ресурс] URL: <http://logic.pdmi.ras.ru/~sergey/teaching/asr/notes-06-features.pdf> (дата обращения: 5.05.2015).

42. Обнаружение радиосигналов [Текст] / П.С. Акимов, Ф.Ф. Евстратов, С.И. Захаров и др.; Под. ред. А.А. Колосова. – М.: Радио и связь, 1989. – 288 с.: ил.

43. Оппенгейм А. Цифровая обработка сигналов [Текст] / А. Оппенгейм, Р. Шафер. – М.: Техносфера, 2006. – 856 с.

44. Первушин, Е.А. Обзор основных методов распознавания дикторов / Е.Н. Первушин // Математические структуры и моделирование. – 2011. - вып.24. – С.41-54.
45. Половикова, О.Н. Использование евклидова и манхэттенского расстояний в качестве меры близости для решения задачи классификации [Текст] / О.Н. Половикова, В.В. Фокина // Известия АлтГУ. – 2010. – №1-1. С.101-102
46. Применение цифровой обработки сигналов [Текст] / под ред. Э. Оппенгейма. М.: Мир, 1980. – 556 с.
47. Рабинер Л. Р., Шафер Р.В. Цифровая обработка речевых сигналов = Digital processing of speech signals[текст]/ Л.Р. Рабинер, Р.В. Шафер.;Пер. с англ.М.В. Назарова, Ю.Н. Прохорова; Под ред. М.В.Назарова, Ю.Н. Прохорова. – Москва: Радио и связь, 1981. – 496с.:ил.
48. Рамишвили, Г.С. Автоматическое опознавание говорящего по голосу[Текст]/Г.С.Рамишвили. – М.: Радио и связь, 1981. - 224 с.
49. Рандалл, Р. Б. Частотный анализ [Текст] / Р. Б. Рандалл. – Глоструп, Дания: К. Ларсен и сын А/О, 1989. – 389 с.
50. Раушер К. Основы спектрального анализа [Текст]: Пер. с англ. С. М. Смольского / Под ред. Ю. А. Гребенко - М.: Горячая линия-Телеком, 2006. – 225 с.
51. Романюк Ю.А. Дискретное преобразование Фурье в цифровом спектральном анализе. Учебное пособие [Текст] / Ю.А. Романюк – М.: МФТИ, 2007. – 120 с.
52. Сергиенко А.Б. Цифровая обработка сигналов: учебное пособие. — 3-е изд.— М.: БХВ-Петербург, 2011. — 768 с.
53. Смит С. Цифровая обработка сигналов: практическое руководство для инженеров и научных работников [Текст] / Стивен Смит ; пер. с англ. А. Ю. Линовича, С. В. Витязева. - Москва : Додэка-XXI, 2008. – 718 с.

54. Солонина, А.И. Основы цифровой обработки сигналов [Текст]/ А.И. Солонина, Д.А. Улахович, С.М. Арбузов, Е.Б.Соловьева. – СПб.: БХВ-Петербург, 2005. – 768с.: ил.
55. Сорокин, В.Н. Верификация диктора по спектрально - временным параметрам речевого сигнала / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы, Том 10. – 2010. - №2. – С.87 –104.
56. Сорокин, В.Н. Распознавание личности по голосу: аналитический обзор / В.Н. Сорокин, В.В.Вьюгин, А.А.Тананыкин // Информационные процессы, Том 12. – 2012. - №1. – С.1 –30.
57. Сорокин В.Н. Сегментация речи на кардинальные элементы [Текст] / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы, 2006, Т. 6, №3, с. 177-207.
58. Сюзев В.В. Основы теории цифровой обработки сигналов [Текст] / В.А. Сюзев – М.: РТСофт, 2014. – 752 с.
59. Трахтман А.М. Введение в обобщённую спектральную теорию сигналов [Текст] / А.М. Трахтман– М.: Изд-во «Советское радио», 1972. – 352 с.
60. Трахтман, А.М. Основы теории дискретных сигналов на конечных интервалах [Текст]/ А.М. Трахтман, В.А. Трахтман.- М.: «Сов. радио», 1975. – 208 с.
61. Ту, Дж., Гонсалес, Р. Принципы распознавания образов / Дж. Ту, Р. Гонсалес; Пер. с англ. И.Б. Гуревича; Под ред. Ю.И. Журавлева. – М.: Мир, 1978 – 411 с.
62. Фирсова, А.А. Разработка и исследование субполосных методов и алгоритмов сегментации речевых сигналов. Диссертация на соискание учёной степени кандидата технических наук – Белгород, 2013. — 165 с.
63. Фланаган, Джеймс Л. Анализ, синтез и восприятие речи =Speech analysis synthesis and perception [Текст]/ Джеймс Л. Фланаган.;Пер. с англ.

А.А. Пирогова; Под ред. А.А. Пирогова. – М.: Издательство «Связь», 1968. – 360с.: ил.

64. Френкс, Л. Теория сигналов [Текст] / Л. Френкс; Пер. с англ., под ред. Д.Е.Вакмана. – М.: Сов. радио, 1974 – 344 с.

65. Фу, К. Структурные методы в распознавании образов / К. Фу; Пер. с англ. Н.В. Завалишина, С.В. Петрова, Р.Л. Шейнина; Под ред. М.А. Айзермана. – М.: Мир, 1977 – 319 с.

66. Фукунага К. Введение в статистическую теорию распознавания образов [Текст] / К. Фукунага; пер. с англ. – М.: Наука, 1979. – 368 с.

67. Цыпкин, Я.З. Информационная теория идентификации [Текст] / Я.З. Цыпкин. – М.: Наука, 1995. – 336 с.

68. Харкевич А.А. Спектры и анализ [Текст] Изд. 5-е / А.А. Харкевич. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 240 с.

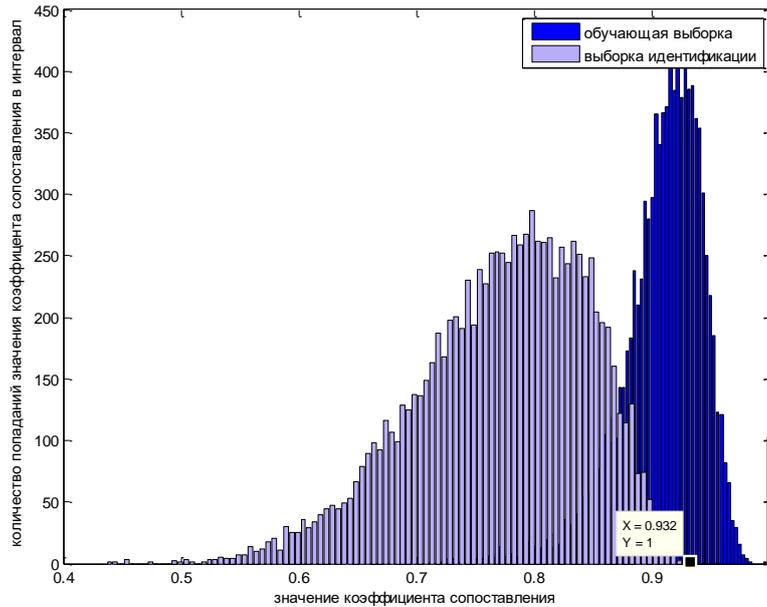
69. Шелухин О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с.

70. Lu X., Dang J. (2007). An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification. *Speech Communication*, v.50, N4, 312–322.

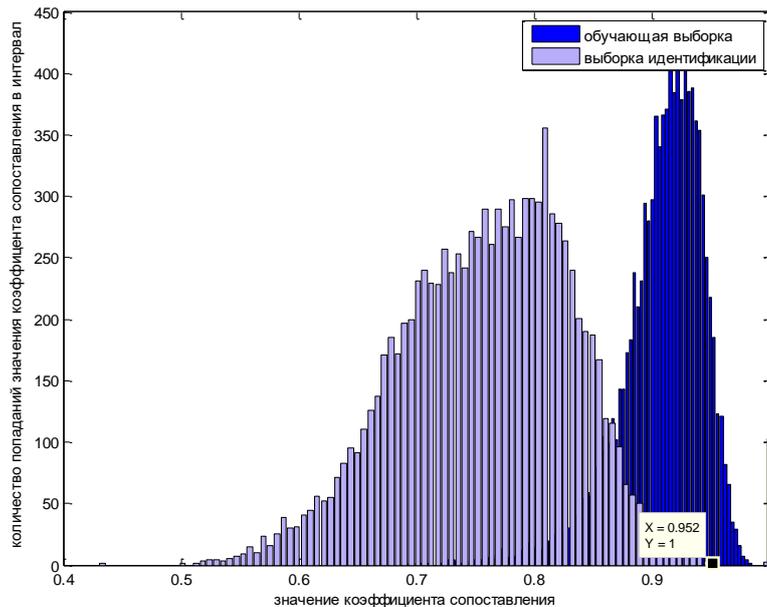
71. В.Г.Потемкин. Справочник по MATLAB [Электронный ресурс] // QRZ.RU: центр компетенций MathWorks. Дата обновления: 15.05.2017. URL: <http://matlab.exponenta.ru/ml/book2/.html> (дата обращения: 10.05.2017).

**ПРИЛОЖЕНИЕ А**

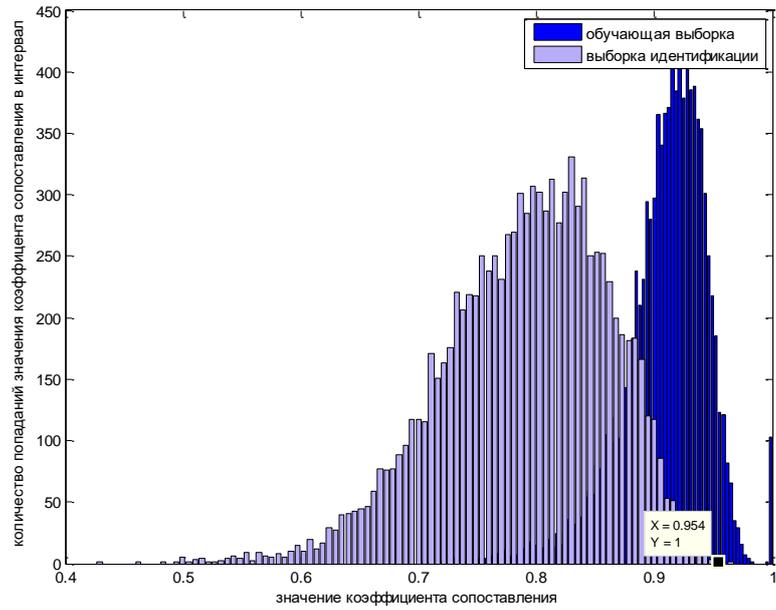
Определение крайних пороговые значений для метода субполосной идентификация дикторов, где качестве признаков используется распределение энергии сигнала по частотным интервалам.



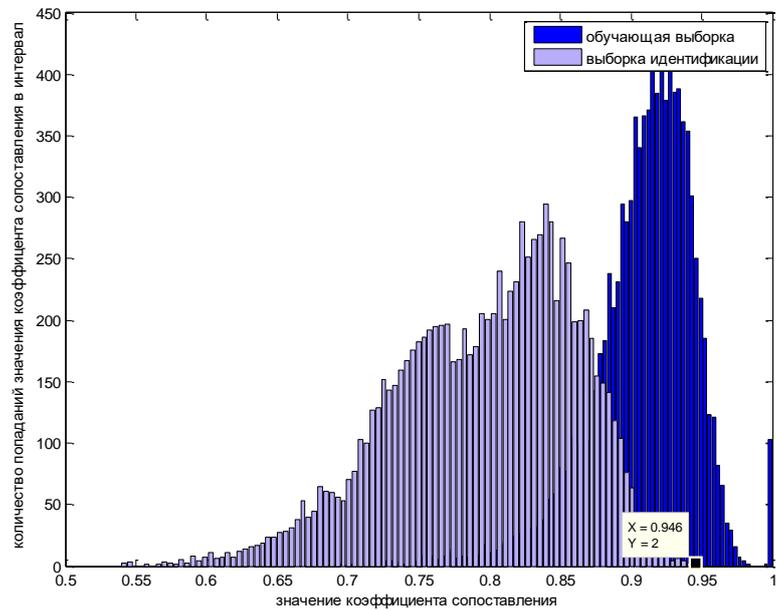
**Рисунок А.1 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 2**



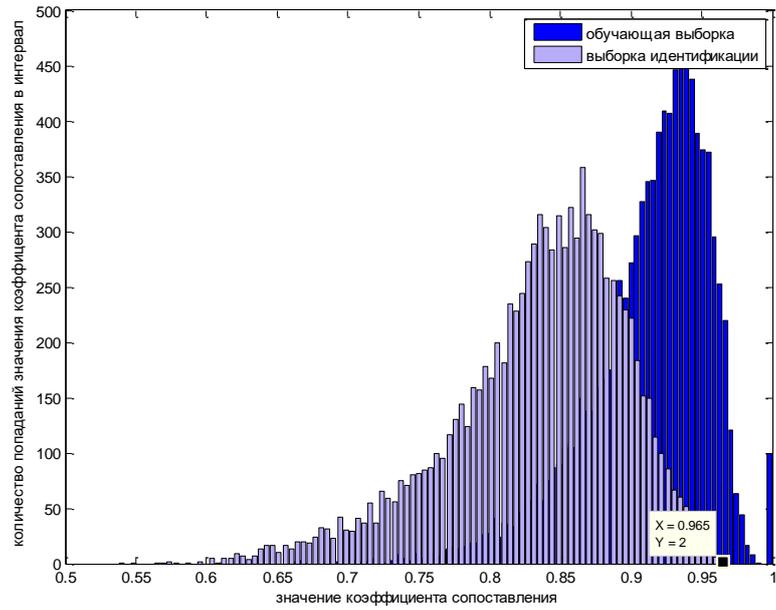
**Рисунок А.2 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 2**



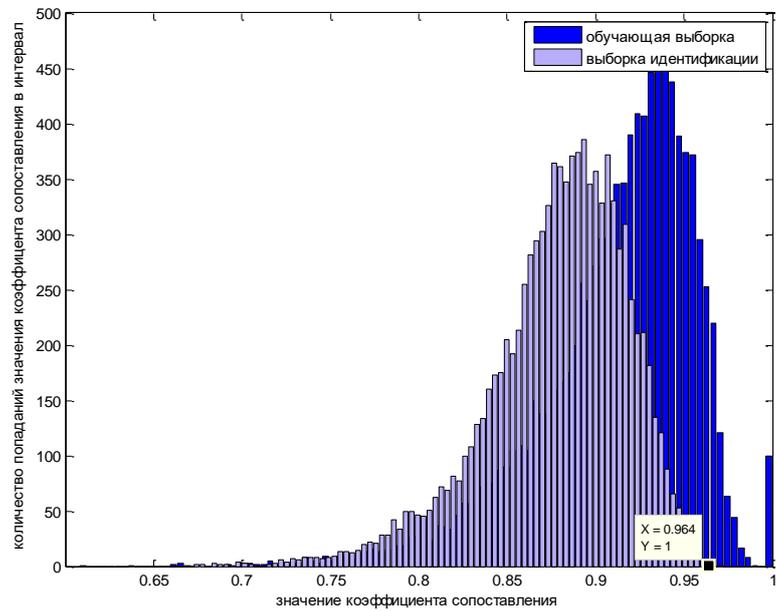
**Рисунок А.3 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 2**



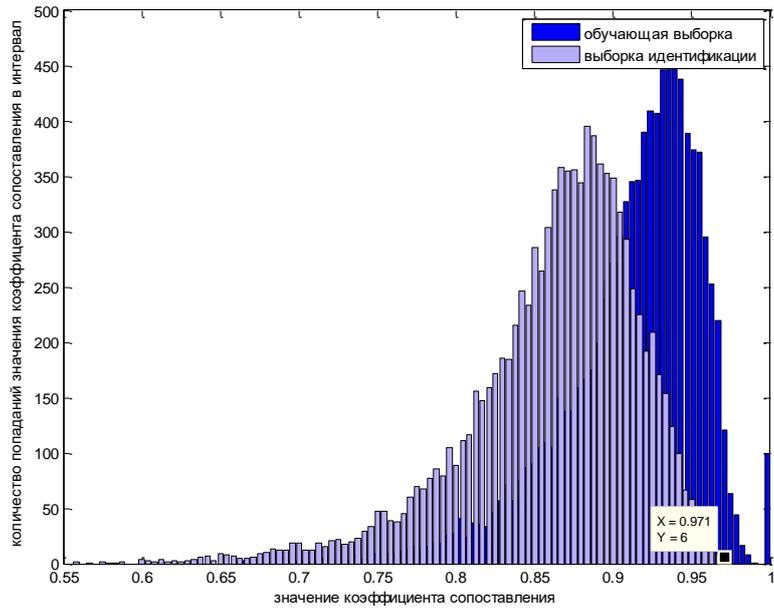
**Рисунок А.4 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 2**



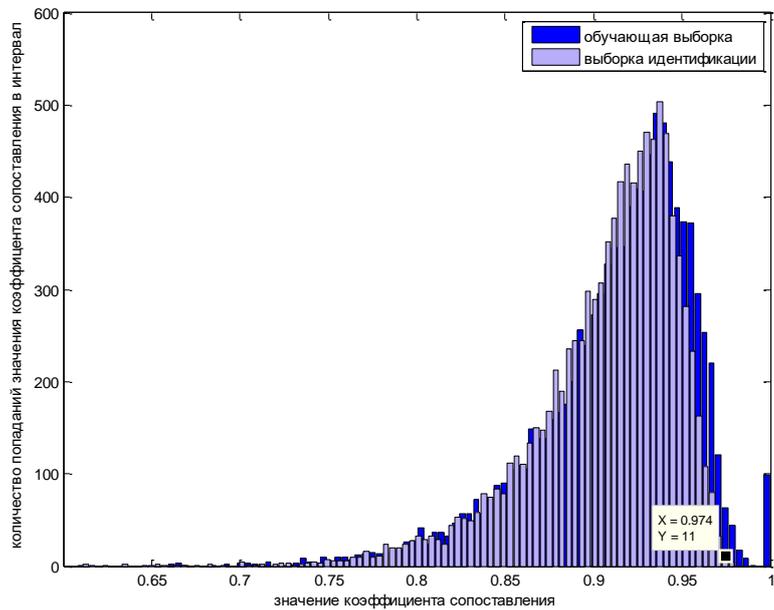
**Рисунок А.5 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 3**



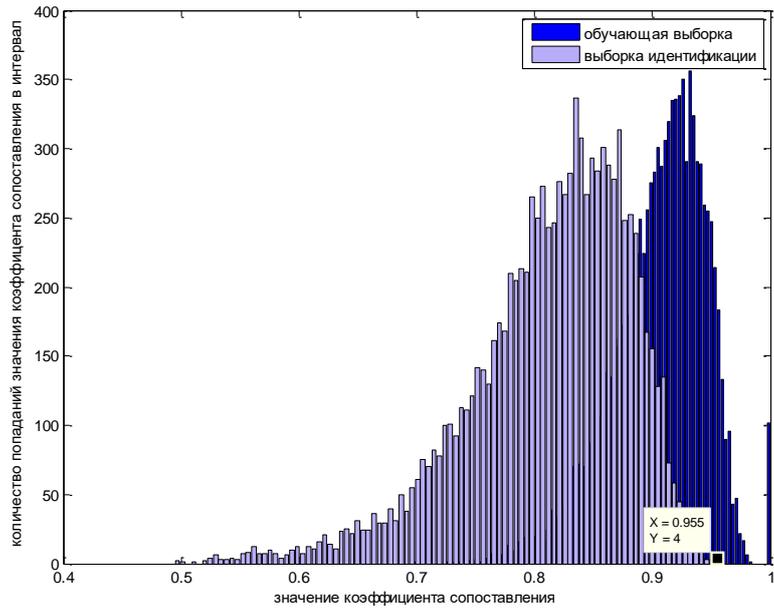
**Рисунок А.6 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 3**



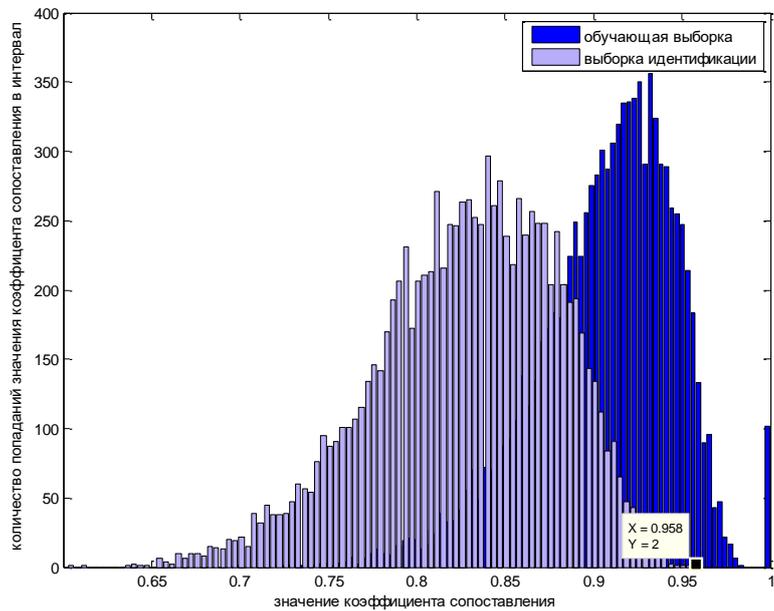
**Рисунок А.7 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 3**



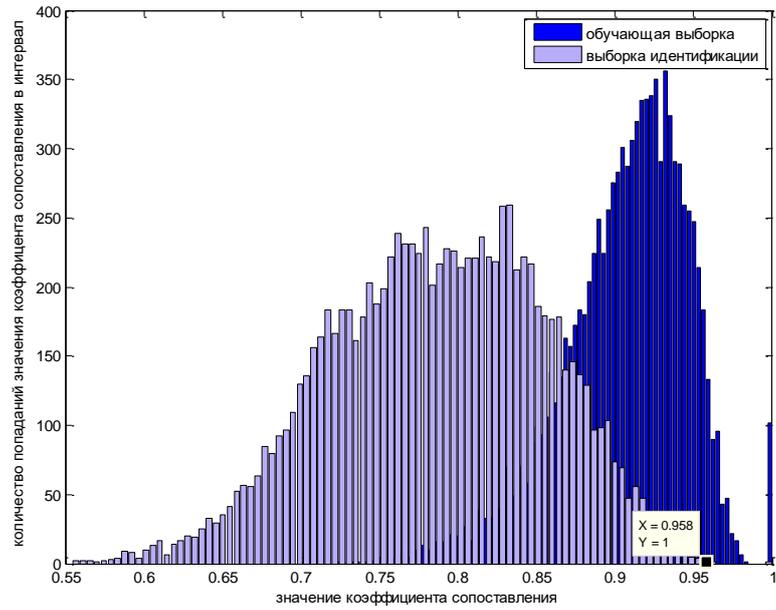
**Рисунок А.8 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 3**



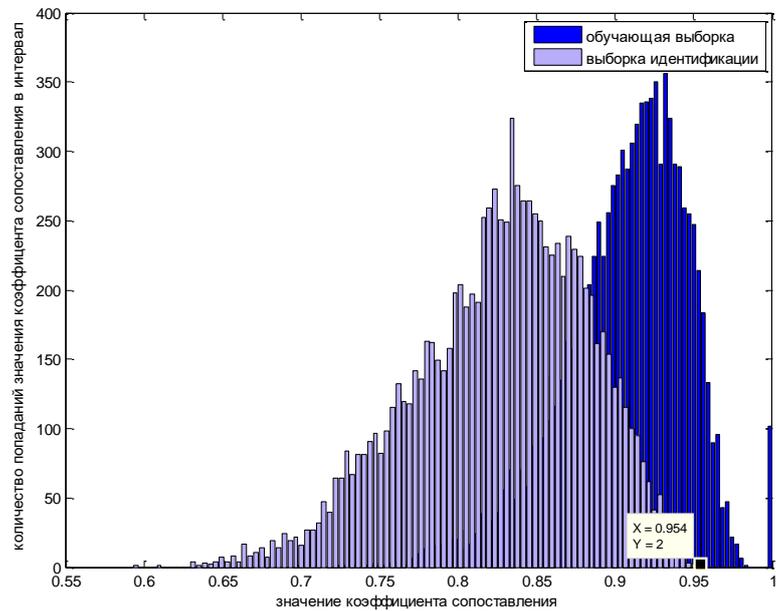
**Рисунок А.9 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 4**



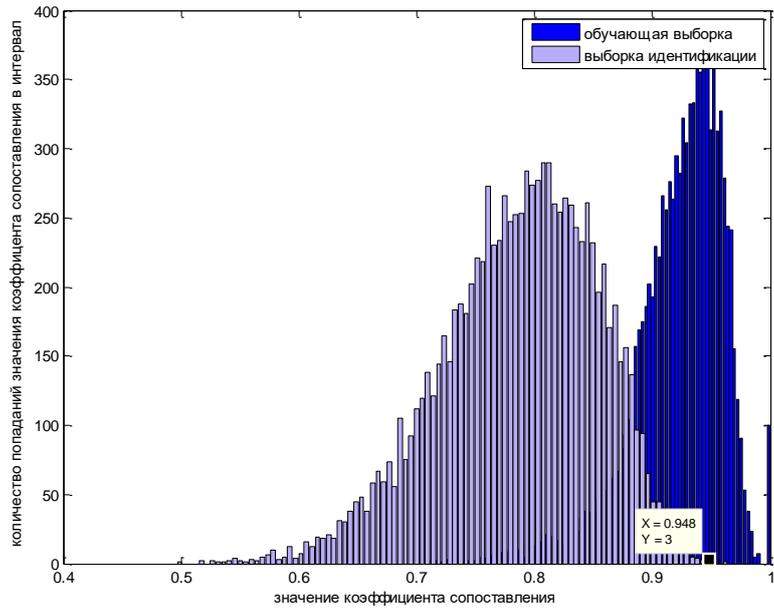
**Рисунок А.10 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 4**



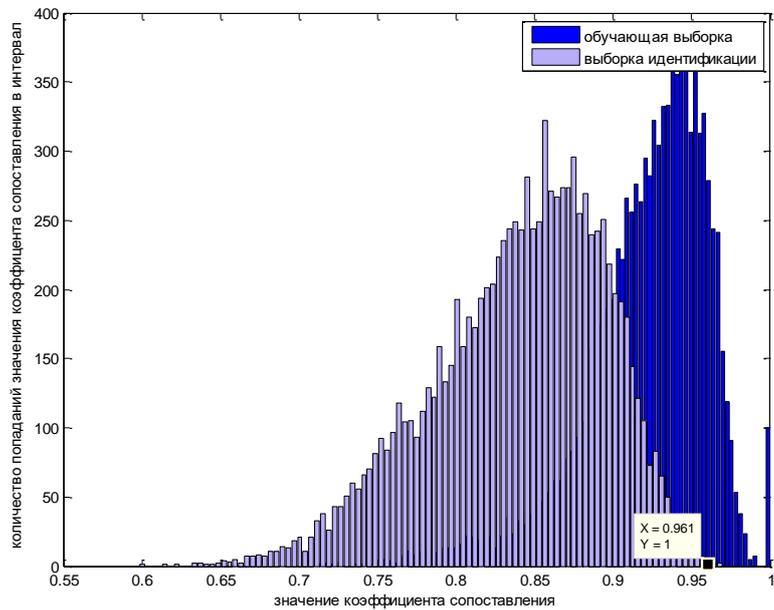
**Рисунок А.11 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 4**



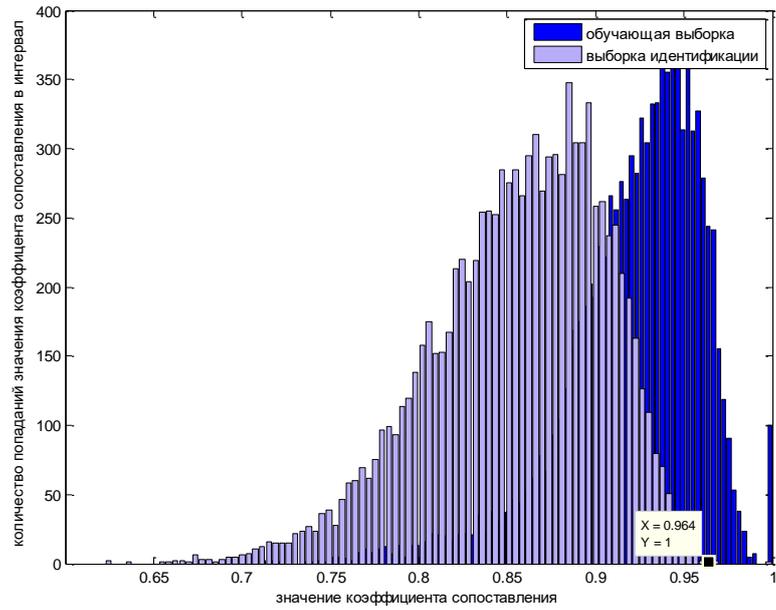
**Рисунок А.12 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 4**



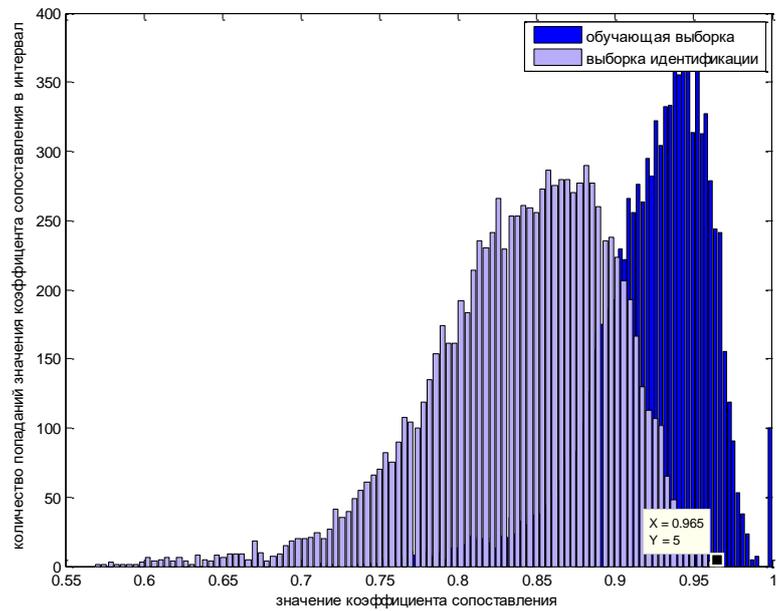
**Рисунок А.13 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 5**



**Рисунок А.14 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 5**



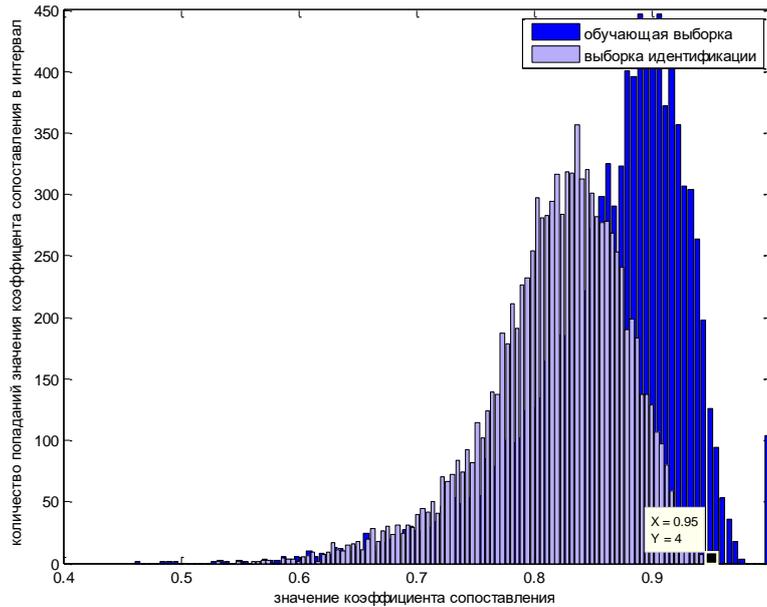
**Рисунок А.15 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 5**



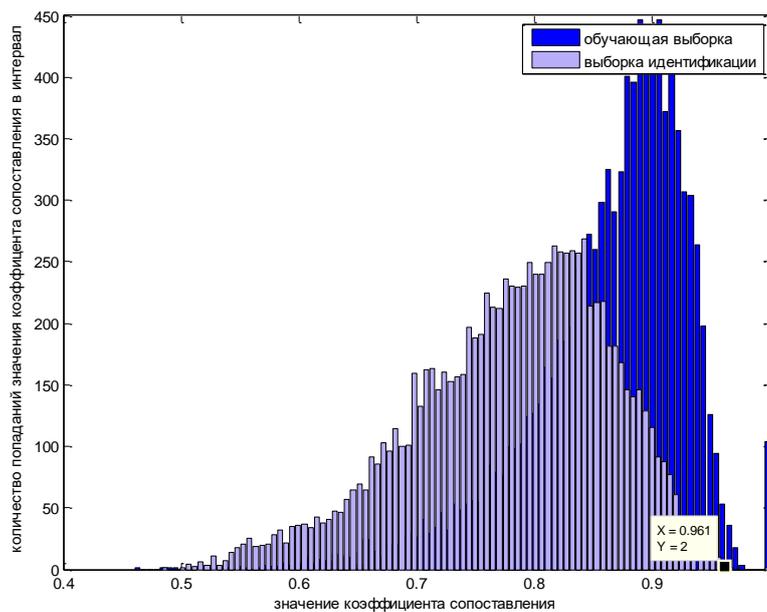
**Рисунок А.16 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 5**

## ПРИЛОЖЕНИЕ Б

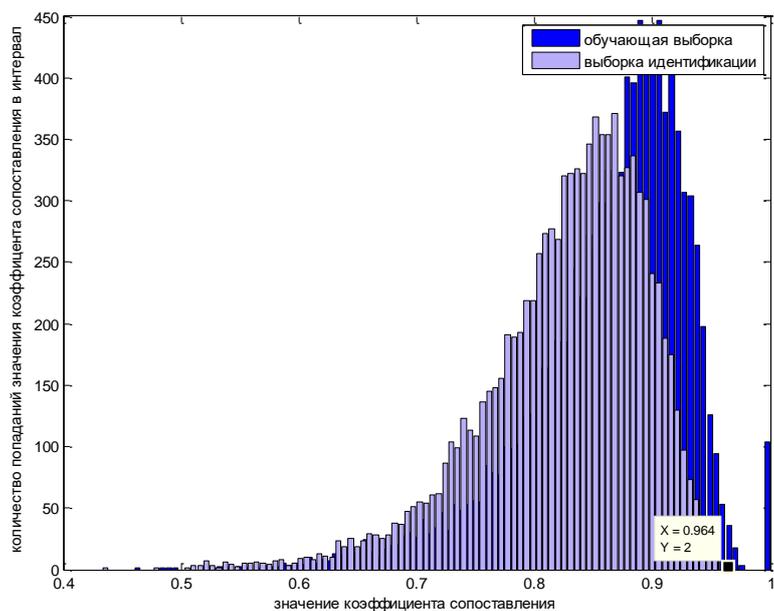
Определение крайних пороговых значений для метода субполосной идентификация дикторов, где качестве признаков используется распределение долей энергии информационных частотных интервалов.



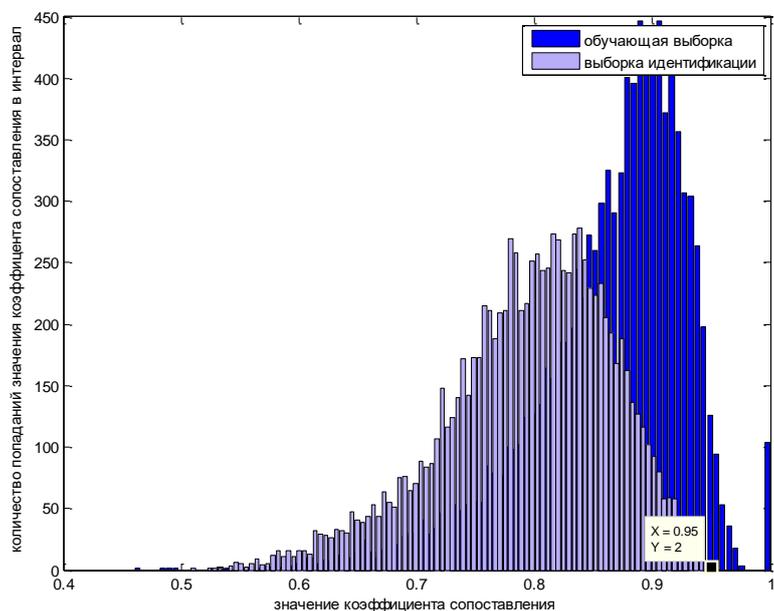
**Рисунок Б.1 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 1**



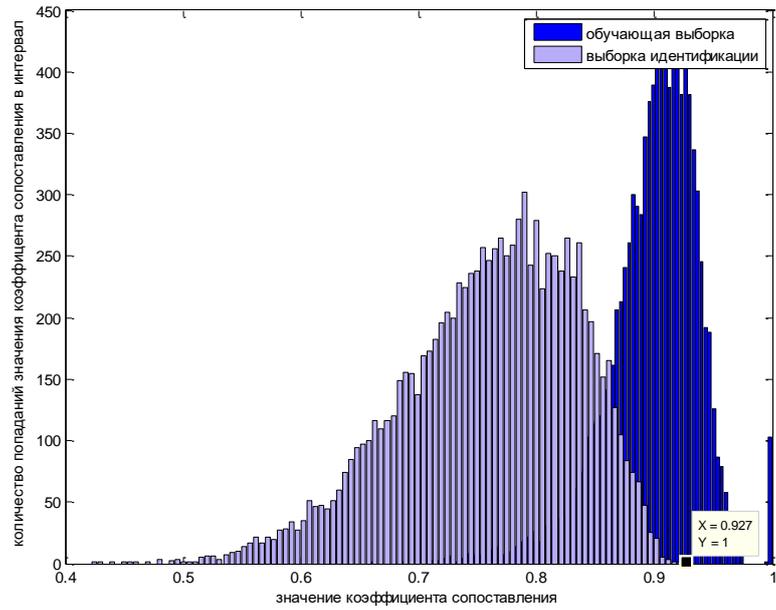
**Рисунок Б.2 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 1**



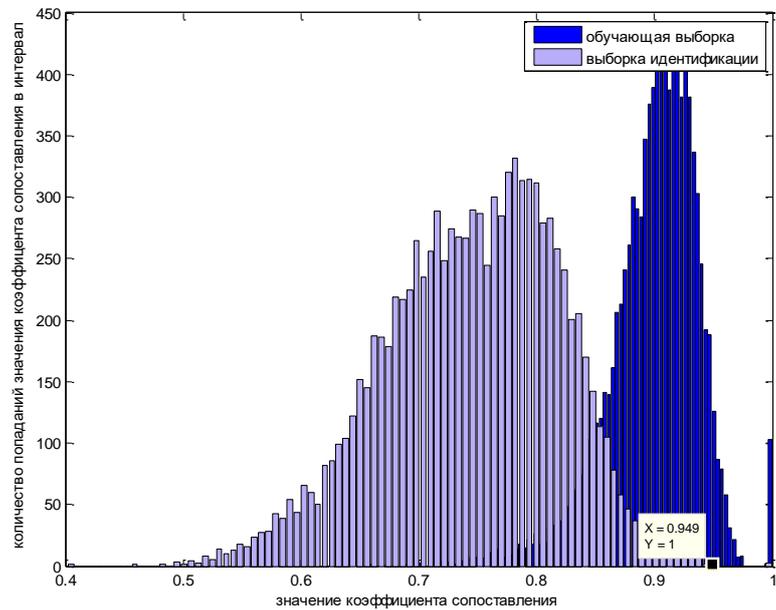
**Рисунок Б.3 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 1**



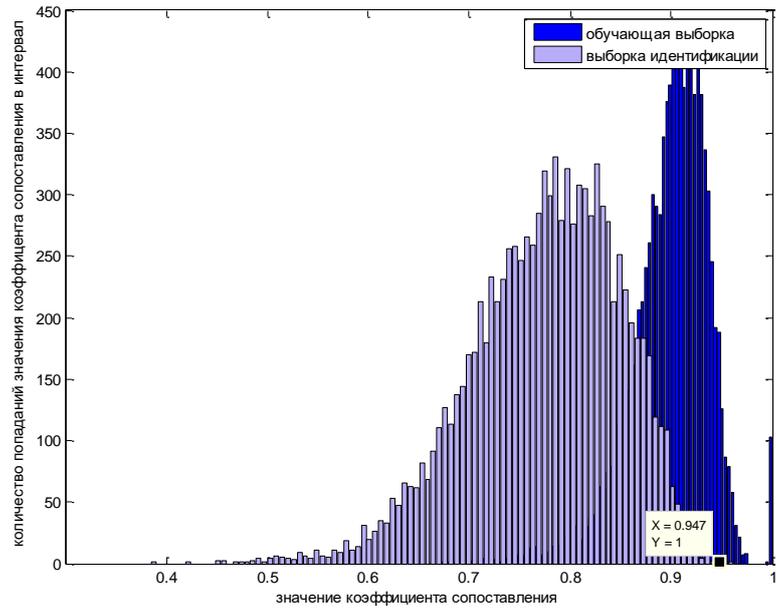
**Рисунок Б.4 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 1**



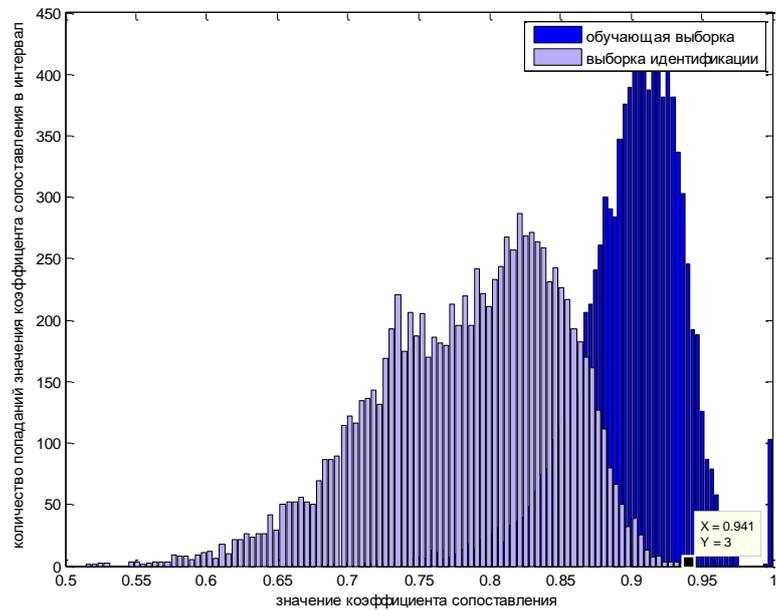
**Рисунок Б.5 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 2**



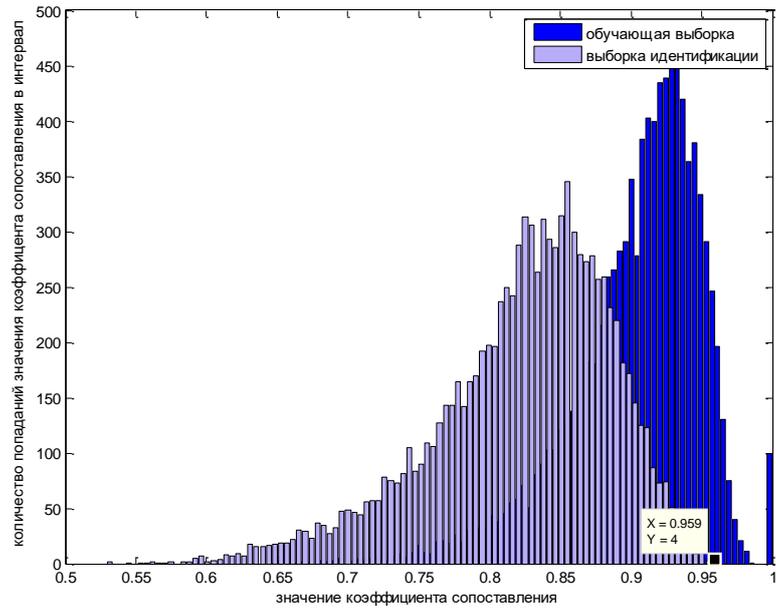
**Рисунок Б.6 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 2**



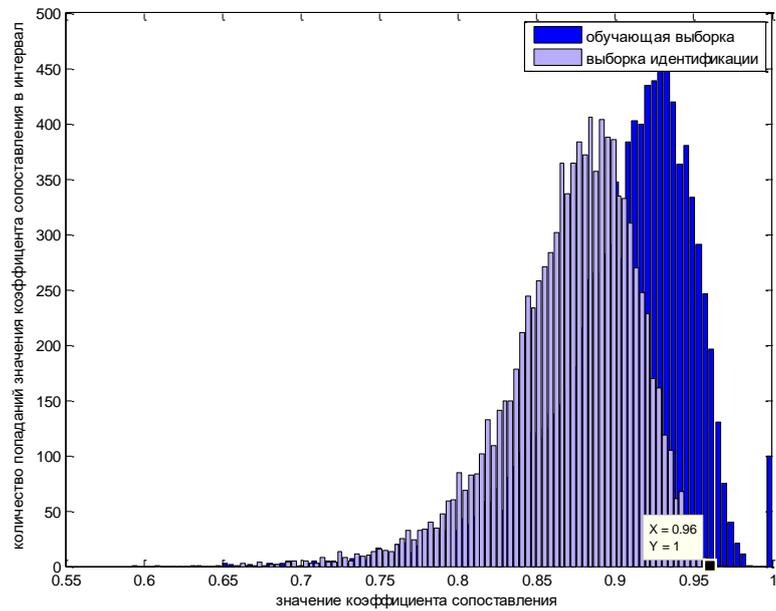
**Рисунок Б.7 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 2**



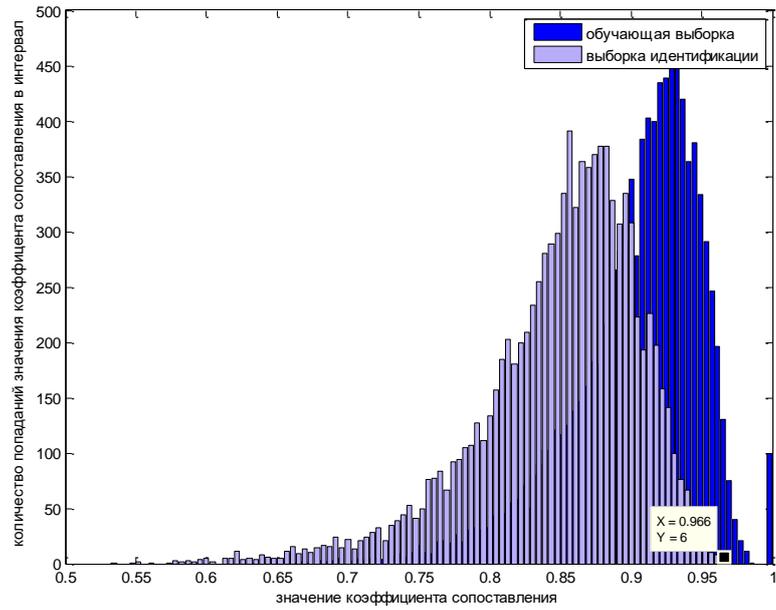
**Рисунок Б.8 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 2**



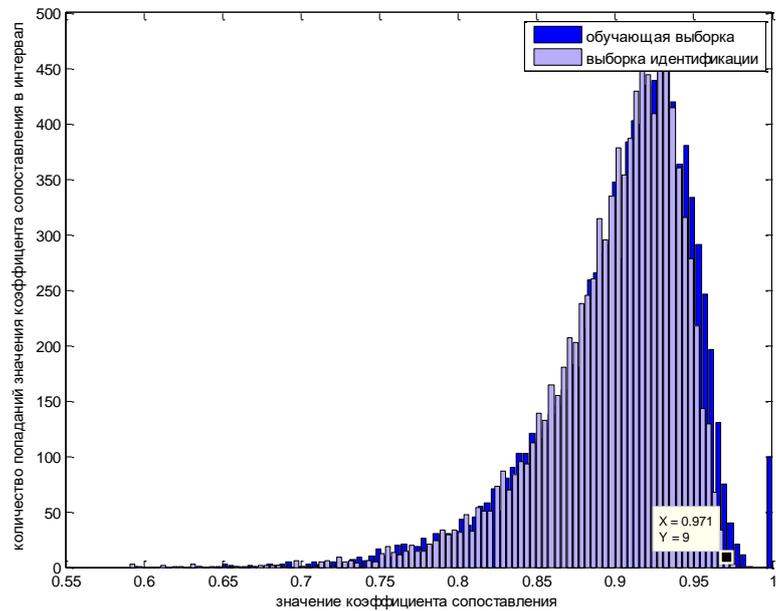
**Рисунок Б.9 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 3**



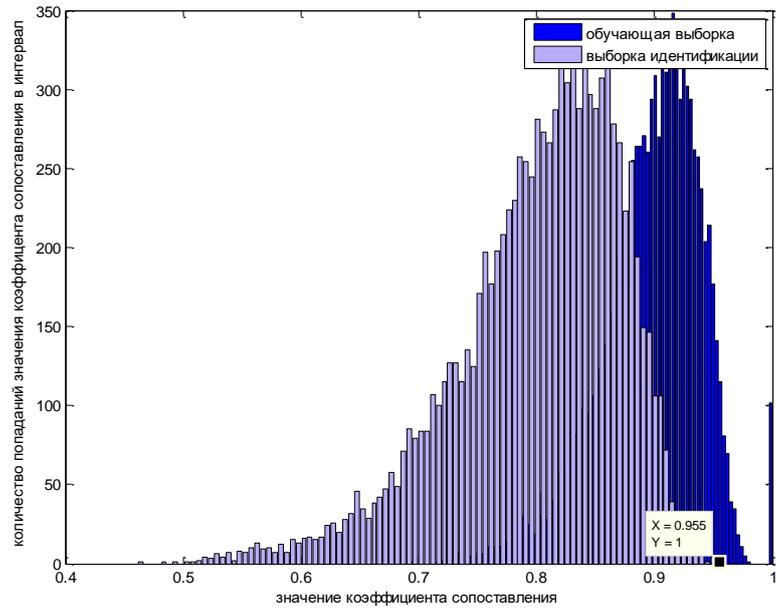
**Рисунок Б.10 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 3**



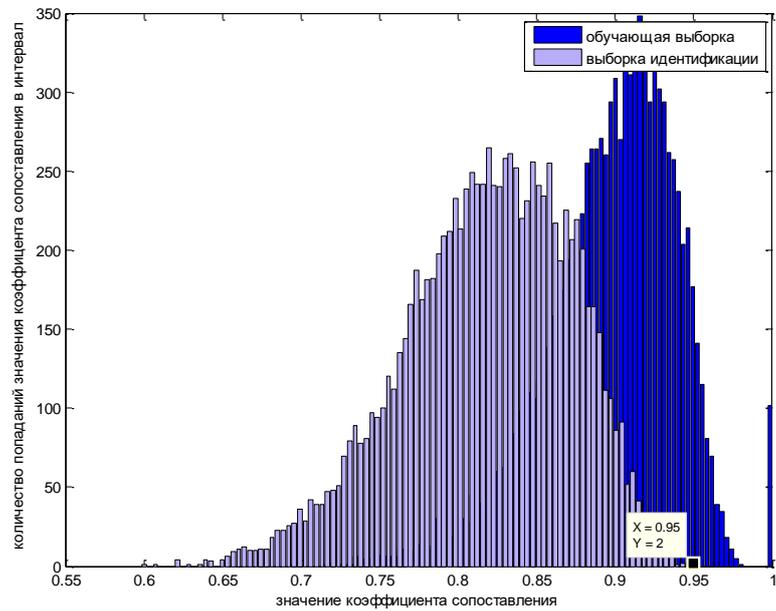
**Рисунок Б.11 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 3**



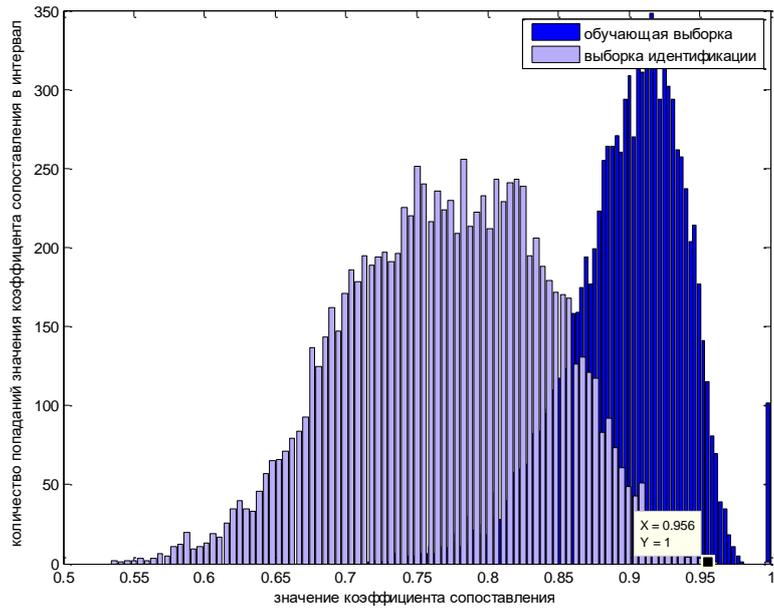
**Рисунок Б.12 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 3**



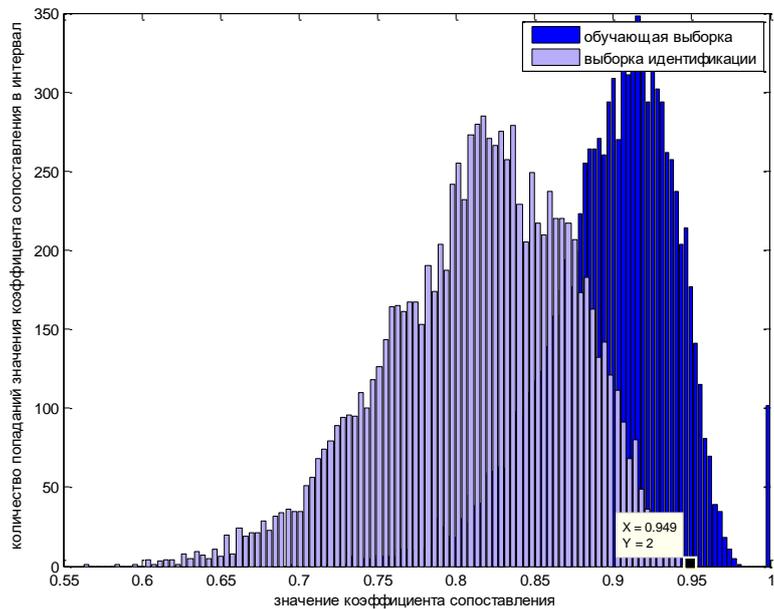
**Рисунок Б.13 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 4**



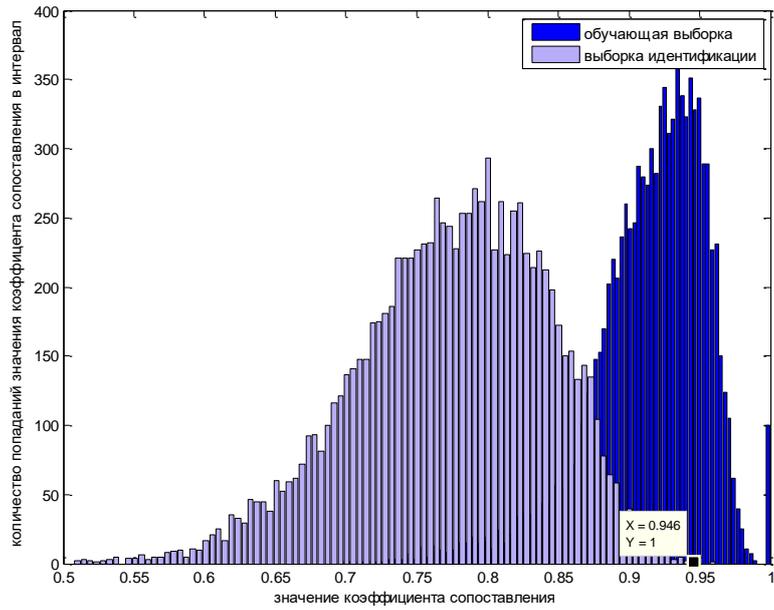
**Рисунок Б.14 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 4**



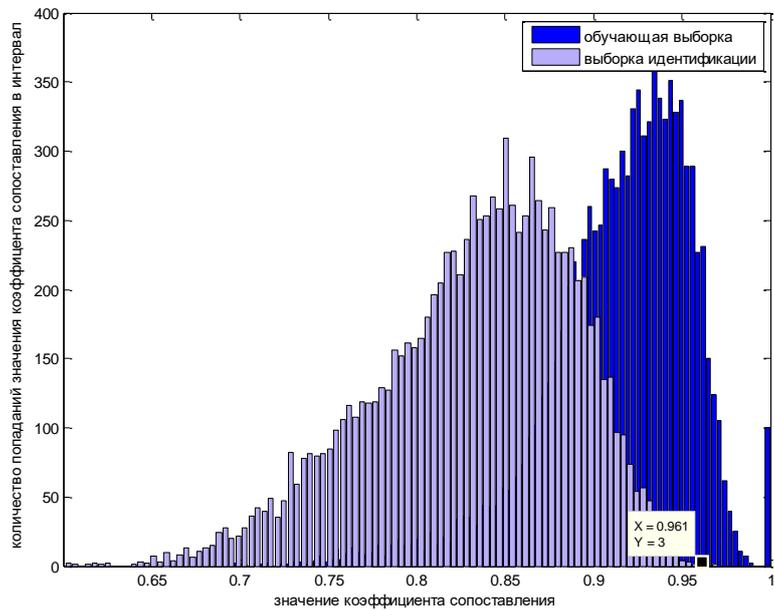
**Рисунок Б.15 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 4**



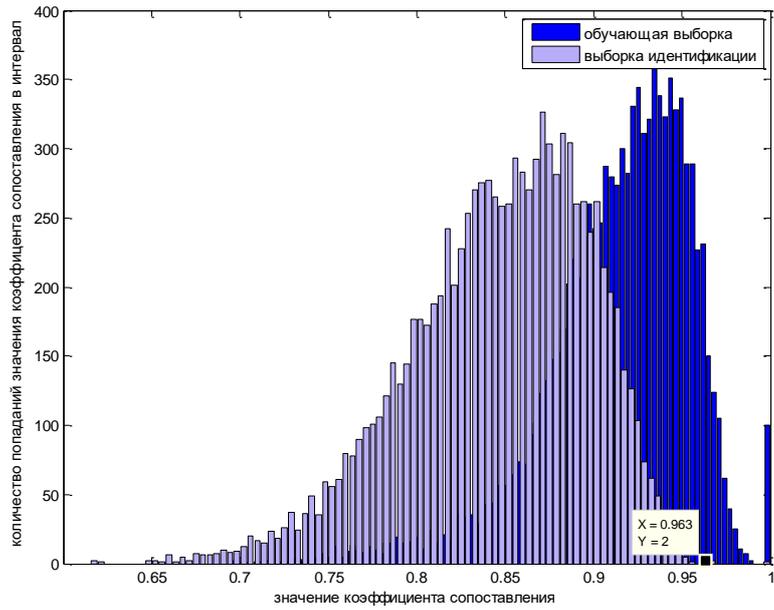
**Рисунок Б.16 – Определение крайнего значения порога для принятия решения при идентификации Диктора 5 как Диктора 4**



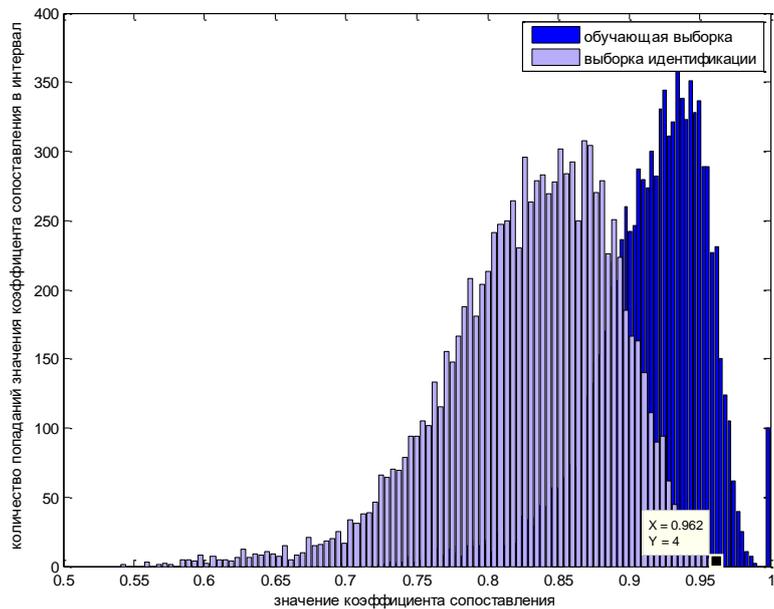
**Рисунок Б.17 – Определение крайнего значения порога для принятия решения при идентификации Диктора 1 как Диктора 5**



**Рисунок Б.18 – Определение крайнего значения порога для принятия решения при идентификации Диктора 2 как Диктора 5**



**Рисунок Б.19 – Определение крайнего значения порога для принятия решения при идентификации Диктора 3 как Диктора 5**



**Рисунок Б.20 – Определение крайнего значения порога для принятия решения при идентификации Диктора 4 как Диктора 5**

Магистерская диссертация выполнена мной совершенно самостоятельно. Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

« \_\_\_ » \_\_\_\_\_ г.

---

*(подпись)*

---

*(Ф.И.О.)*