Санкт-Петербургский государственный университет
**Кафедра математического моделирования энергетических систем**

# Куприянова Елена Геннадьевна

**Выпускная квалификационная работа бакалавра**

# Оценивание частоты основного тона речевого сигнала на основе корреляционных методов

Направление 010400
Прикладная математика и информатика

Научный руководитель,
д–р физ.-мат. наук,
доцент
Михеев С. Е.

Санкт-Петербург
2017

# Содержание

# Введение

Данная работа посвящена проблеме оценивания периода основного тона (ОТ) звукового сигнала.

Речь есть средство обмена информацией между людьми, в отличие от сигналов, получаемых на выходе технических систем, для речевого сигнала характерна большая вариативность даже при передаче совершенно идентичных сообщений. Акустический речевой сигнал, в отличие от письменной речи, переносит огромное количество дополнительной информации, связанной со смысловым значением сообщения (семантика), с индивидуальностью голоса диктора, с эмоциональным характером и стилем высказывания, типом речевого сообщения (монолог, диалог и т.п.), с окружающей обстановкой, состоянием голосового аппарата, половой принадлежностью, возрастом, ростом и весом диктора. Одним из важнейших параметров речевого сигнала является основной тон, содержащий информацию об интонационной структуре произнесения, особенности голоса диктора и его эмоциональном состоянии. Оценивание частоты (или периода) основного тона является одной из наиболее важных задач в обработке речи. Выделители основного тона используются в вокодерах [1], системах распознавания и идентификации дикторов [2], в устройствах, предназначенных для глухих [1,2]. Поскольку задача выделения основного тона очень важна, существует ряд способов ее решения [1]. Все они обладают ограничениями и наиболее естественным будет признать, что в настоящее время отсутствует метод выделения основного тона, обеспечивающий удовлетворительные результаты для различных дикторов, в разных областях применения и условиях эксплуатации.

Для определения основного тона оцифрованных звуковых сигналов в работе применены модификации корреляционного анализа

# Постановка задачи

Цель работы — Получить работоспособный метод определения основного тона с достаточно высоким быстродействием.

Для решения поставленной цели предложены следующие задачи:

Задача 1: Провести сравнительный анализ существующих методов выделения основного тона, найти их модификации, повышающие качество и быстродействие.

Задача 2: На основе имеющихся алгоритмов создать программный комплекс, позволяющий за приемлемое время решать вопросы определения основного тона в файлах типа WAV с оцифрованным звуком.

# Глава 1. Акустика. Понятие основного тона для музыкальных и речевых колебаний.

Звук – это колебания, то есть периодическое механическое возмущение в упругих средах – газообразных, жидких и твердых. Такое возмущение, представляющее собой некоторое физическое изменение в среде (например, изменение плотности или давления воздуха), распространяется в нем в виде звуковой волны. Область физики, рассматривающая вопросы возникновения, распространения, приема и обработки звуковых волн, называется акустикой. Звук может быть неслышимым, если его частота лежит за пределами чувствительности человеческого уха, или он распространяется в такой среде, как твердое тело, которая не может иметь прямого контакта с ухом, или же энергия колебаний частиц среды, переносящей звуковые волны, быстро рассеивается в среде. Таким образом, обычный для нас процесс восприятия звука – лишь одна задача акустики. В этой области основным тоном называется звук, который создаёт акустическая система, когда колеблется с низшей возможной для неё частотой. За единицу измерения частоты колебаний принят Герц, равный одному колебанию в секунду. Термин «основной тон» применяют также для обозначения составляющей с наинизшей частотой при разложении сложного периодического колебания на более простые в базисе ортогональных функций (результат спектрального анализа звука). В качестве базиса используют синусоидальные функции, поскольку они определены при всех значениях времени, ортогональны и составляют полный набор при кратном уменьшении периода. В качестве разложения обычно используются преобразование Фурье, разложение по функциям Уолша, вейвлет-преобразование и др [4]. Обертоны - это все составляющие спектра выше низшей частоты (основной тон), а обертоны, частоты которых относятся к частоте основного тона как целые числа, называются гармониками, причем основной тон считается первой гармоникой. Совокупность обертонов, составляющих сложный звук, называют спектром этого звука.

Среди слышимых звуков следует особо выделить речевые звуки (из которых состоит устная речь) и музыкальные звуки (из которых состоит музыка). Рассмотрим их подробнее.

## 1.1. Музыкальная акустика

Музыкальный звук – это звук с определённой точной высотой, имеющий интерпретацию с позиции музыкальной логики: по местоположению и значению в звуковой системе, в звукоряде, в созвучиях (интервалах, конкордах, аккордах и т. д.). Музыкальный основной тон – это периодический звук, то есть колебания, которые снова и снова повторяются через определенный период. Периодический звук можно представить в виде суммы колебаний с частотами, кратными частоте основного тона. В музыке обычно используются звуки, основная частота которых лежит от субконтроктавы до 5-й октавы. Так на рис.1 звуки стандартной 88-клавишной клавиатуры фортепиано укладываются в диапазон от ноты ля субконтроктавы (27,5 Гц) до ноты до 5-й октавы (4186,0 Гц).


Рис. 1: Клавиатура фортепиано по октавам

Однако музыкальный звук обычно состоит не только из чистого звука основной частоты, но и из примешанных к нему гармоник (звуков с частотами, кратными основной частоте), а иногда и из шумовых компонент в широком диапазоне частот. Иначе говоря, музыкальный звук представляет из себя смесь основного тона и обертонов. Обертоны музыкальных звуков лежат во всём доступном для слуха диапазоне частот (в среднем от 20 Гц до 20 кГц). Их наличие обусловлено сложной картиной колебаний звучащего тела (струны, столба воздуха, мембраны и т. д.): частоты обертонов соответствуют частотам колебания его частей. Обертоны бывают гармоническими и негармоническими. Частоты гармонических обертонов кратны частоте основного тона (гармонические обертоны вместе с основным тоном также называются гармониками), в реальных физических ситуациях (например, при колебаниях массивной и жесткой струны или при использовании не настроенного инструмента) частоты обертонов могут заметно отклоняться от величин, кратных частоте основного тона — такие обертоны

называются негармоническими. Присутствие негармонических обертонов в колебаниях струн музыкальных инструментов приводит к феномену неточного равенства между рассчитанными частотами равномерно темперированного строя, в котором каждая октава делится на математически равные интервалы, в наиболее типичном случае — на двенадцать полутонов, и реальными частотами правильно настроенного фортепиано. Ввиду исключительной важности для музыки именно гармонических обертонов (и относительной малозначимости негармонических) вместо «гармонический обертон» в музыкально-теоретической литературе часто пишут «обертон» без каких-либо уточнений. С точки зрения определения частоты основного тона музыкальный звук примечателен тем, что обертона не искажают основной тон, следовательно, отдельно первая гармоника будет в некотором приближении повторять первоначальную мелодию, что является контрольным этапом тестирования программы впоследствии.

## 1.2. Речевая акустика

Речь формируется при прохождении выталкиваемого легкими потока воздуха через голосовые связки и голосовой тракт. Один из способов описания речи заключается в представлении ее в виде сигнала, то есть акустического колебания, что является целесообразным на практике. Речевое общение начинается с того, что в мозгу диктора возникает в абстрактной форме некоторое сообщение. В процессе речеобразования это сообщение преобразуется в акустическое речевое колебание. Информация, содержащаяся в сообщении, представлена в акустическом колебании весьма сложным образом. Сообщение сначала преобразуется в последовательности нервных импульсов, управляющих артикуляторным аппаратом на рис. 2 (т. е. перемещением языка, губ, голосовых связок и т. д.). В результате воздействия нервных импульсов артикуляторный аппарат приходит в движение, результатом которого является акустическое речевое колебание, несущее информацию об исходном сообщении. В системах речевой связи сигнал передается, хранится и обрабатывается различными способами. Задачи техники обусловливают применение различных форм представления речевого сигнала, которое должно быть таким, чтобы его информационное содержание легко воспринималось автоматически с помощью машины или при прослушивании человеком. Обработка сигнала предполагает в первую оче-

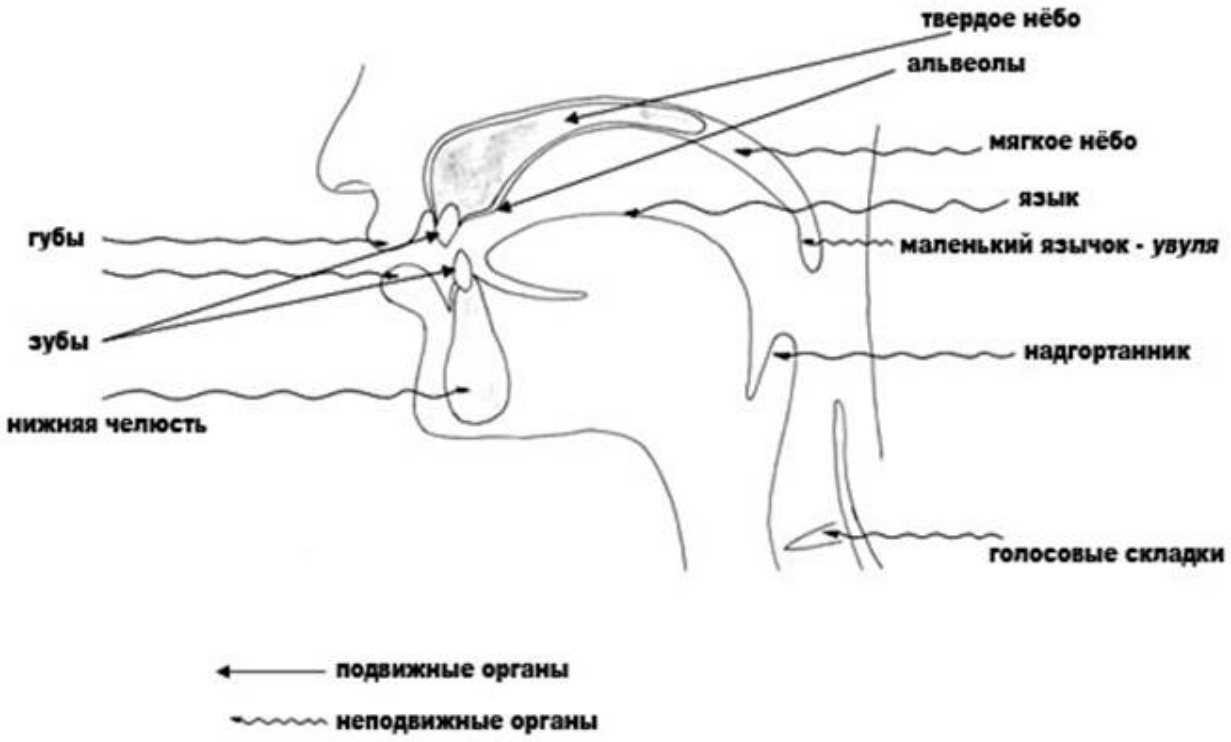


Рис. 2: Строение речевого аппарата

редь формирование совокупности физических параметров, определяющих процесс восприятия речи (на основе некоторой модели) с последующим преобразованием полученного представления в требуемую форму. Последним шагом в процессе обработки является выделение и использование информационного содержания сигнала.

В частности, таким информационным содержанием сигнала являются признаки индивидуальности диктора, среди которых выделяют понятие основного тона. Считается, что основной тон - это одна из акустических характеристик, определяющая восприятие речевой интонации и структуру речевого сообщения. Эта акустическая характеристика связана с частотой открывания и закрывания (или вибрации) голосовых связок. Исходящий из голосового аппарата звук кроме основного тона включает также многочисленные сопутствующие ему тоны. Основной тон и эти дополнительные тоны составляют сложный звук. Голосовой сигнал может быть представлен в виде пульсирующего воздушного потока, где частота повторения импульсов голосового источника носит название частоты основного тона. Если у

диктора она наблюдается постоянной, то речь будет восприниматься как машинная и монотонная. Если связки вибрируют быстро, это соответствует более высокой частоте основного тона. У возбужденного диктора частота основного тона его речи обычно высока. Она может меняться в зависимости от эмоциональной окраски речи, но в достаточно узких пределах. Изменение частоты основного тона называется интонацией. У каждого человека свой диапазон изменения основного тона (обычно он бывает немногим более октавы) и своя интонация. Частота основного тона речи для всех голосов лежит в пределах от 75 (низкий тон низкого мужского голоса) до 500 (высокий тон женского или детского голоса) Гц [5]. Еще больший диапазон изменений частоты основного тона голоса достигается при пении. Типичные средние значения частоты основного тона в речи, определенные на группе говорящих, составляют 132 Гц для мужчин, 223 Гц для женщин и 264 Гц для детей [5].

Величина, обратная частоте основного тона, называется периодом основного тона $P$. Эта величина также является индивидуальной характеристикой диктора. Как правило, он определяется временным интервалом между последовательными циклами раскрытия и закрытия голосовых связок. Несмотря на то, что величина основной частоты (усредненная по некоторому соответствующим образом выбранному временному интервалу) может быть непосредственно оценена по спектру речевого колебания, обычно средний период основного тона $P$ оценивается по некоторому отрезку последовательности отсчетов речевого сигнала. Соответствующая величина основной частоты для анализируемого сегмента тогда определяется как $1/P$.

Следует подчеркнуть, что изменение частоты основного тона во времени имеет сложную структуру. Соседние периоды основного тона, как правило, отличаются по величине друг от друга, и эти различия передают разную информацию. Основная трудность заключается в отсутствии точного определения основного тона. Качественно основной тон есть субъективное свойство, которое позволяет расположить по шкале частот весь диапазон изменений голоса от низкого до самого высокого. Вокализованное возбуждение голосового тракта носит исключительно квазипериодический характер. Сигнал, создаваемый колебаниями голосовых связок, изменяются не только по амплитуде и длительности периода, но также и по фор-

ме. Точно указать, какие интервалы речевого сигнала или даже сигналы возбуждения от голосовых связок должны быть выбраны в качестве измеряемых периодов, не представляется возможным. Не установлена также достаточно четкая связь между измеренными интервалами и воспринимаемым основным тоном. Однако существуют некоторые пути решения этой проблемы.

# Глава 2. Обзор существующих алгоритмов оценивания частоты основного тона речи

В большинстве методов выделения основного тона в качестве объекта измерения используются интервалы между соседними импульсами, появляющимися с частотой колебания голосовых связок. За последнее время было предложено несколько подходов, которые по характеристикам превосходят традиционные. Какие-то методы обладают большей точностью, а какие-то большей устойчивостью к шумам. Все алгоритмы выделения частоты основного тона можно разделить на алгоритмы, основанные на: амплитудных методах, частотном анализе, учете корреляционных свойств речевых сигналов. Далее рассматриваются некоторые алгоритмы для каждой из указанных выше групп.

## 2.1. Амплитудная селекция

Идея состоит в том, что на стационарном участке вокализованного сигнала при незначительном уровне шумов, форма речевого колебания почти так же повторяется на каждом очередном периоде основного тона. Расстояние между максимумами речевого сигнала можно приблизительно считать равным за период основного тона. Основная проблема алгоритмов амплитудной селекции состоит в необходимости подавления ложных локальных максимумов. Этого можно добиться с помощью повышения порогового значения обнаружения в поиске максимумов. Однако при этом увеличивается вероятность ошибочного определения за счет пропуска истинного максимума. Пропуск или потеря максимума может привести к существенным искажениям звука после обработки. Добавление второго канала амплитудной селекции, выделяющего положение минимумов речевого сигнала, повышает надежность определения периода основного тона.

Главным достоинством устройств временной селекции является простота в реализации. Основной недостаток – невысокая точность и неустойчивое определение основного тона даже при относительно небольшом уровне шумов.

## 2.2. Частотная селекция

В спектре звукового сигнала присутствуют пики на частотах, кратных частоте основного тона. Если построить дискретное преобразование Фурье с достаточно малым шагом дискретизации по частоте, то в качестве оценки частоты основного тона можно использовать частоту, соответствующую максимальному значению энергии спектра. Понятия энергии и мощности в теории сигналов не относятся к характеристикам каких-либо физических величин сигналов, а являются их количественными характеристиками, отражающими определенные свойства сигналов и динамику изменения их значений во времени. Для произвольного, в общем случае комплексного, энергия сигнала равна интегралу от мощности по всему интервалу существования или задания сигнала. Мощность по определению равна квадрату функции его модуля, для вещественных сигналов - квадрату функции амплитуд. Поиск максимума энергии спектра следует производить в интервале 80 – 400 Гц. Однако имеет место ситуация, когда в указанной полосе лежит и вторая гармоника основного тона, иногда обладающая большей энергией. В этом случае она будет ошибочно принята за оценку основного тона. Чтобы избежать этого, ищется максимум не спектра $X_n(k)$ , а следующей функции уплотнения спектра

$$P_n(k) = \prod_{r=1}^{R} |X_n(kr)|^2, \tag{1}$$

индекс $n$ указывает на то, что спектр $X_n(k)$ и функция $P_n(k)$ вычисляются в момент времени $n$. Поскольку логарифм монотонно возрастает в области допустимых значений, целевая функция примет следующий вид

$$\tilde{P}_n(k) = \frac{1}{2}\ln(P_n(k)) = \sum_{r=1}^{R} ln(|X_n(kr)|). \tag{2}$$

Эта функция представляет собой сумму $R$ сжатых по частоте в $r$ раз логарифмов спектра мощности. Суть идеи в том, что для истинной частоты основного тона вторая гармоника второго слагаемого сложится с первой гармоникой первого слагаемого, что усилит ее. Аналогично для последующих слагаемых. В результате для вокализованного речевого сигнала будет существовать выраженный пик функции $\tilde{P}_n(k)$ на частоте основного тона. Для невокализованного звука суммирование будет иметь хаотический

характер.

В общем случае оценка значений спектра может оказаться несостоятельной и иметь большие погрешности. Для повышения точности оценки спектральных составляющих, например, нормированной спектральной плотности мощности, применяют методику спектральных окон. Выбор спектрального окна при анализе определяется в результате оптимального выбора между разрешающими способностями по частоте и во времени [6]. Однако применение нелинейного преобразования спектра и окон может вносить большие смещения, что существенно снижает точность оценки.

## 2.3. Корреляционные методы

В основе корреляционных методов определения периода ОТ речевого сигнала заложены принципы оценки среднего значения периода пульсаций квазипериодической корреляционной функции [7]. Корреляционная функция является Фурье-преобразованием энергетического спектра, и положения ее пиков соответствуют расстояниям между равномерно расположенными гармониками спектра. В частном случае вычисляется первый глобальный максимум корреляционной функции [7]. Однако область частот первой форманты (область усиленных частот) достаточно ощутимо влияет на качество работы корреляционного анализа. Важнейшим параметром, характеризующим спектр (распределение энергии или амплитуды по частотам) речевого сигнала, являются форманты, представляющие собой концентрации энергии в ограниченной частотной области. Форманта характеризуется частотой, шириной и амплитудой. За частоту форманты принимают частоту максимальной амплитуды в ее ограниченных пределах. Другими словами, форманта – это некоторый амплитудный всплеск на графике спектра, а его частота – частота пика этого всплеска. Голосовой тракт в силу своих резонансных свойств вносит в формируемый сигнал набор характерных для каждого человека частотных составляющих, называемых формантами. Частоты и полосы этих формант могут управляться изменением формы голосового тракта, например, изменением положения языка. Для выравнивания спектра может быть использована либо обратная фильтрация на основе линейного предсказания, либо разделение сигнала на несколько частотных полос с вычислением корреляционной функции в каждой полосе с нормировкой и суммированием. [3, 8]

Данный подход обеспечивает существенно более высокую достоверность оценки периода основного тона по сравнению с методами амплитудной или частотной селекции. При этом алгоритм обладает значительной вычислительной сложностью. Корреляционные методы оценивания периода основного тона имеют общий недостаток: неустойчивую работу в случае, когда речевой сигнал модулирован по амплитуде. В практике часто речевые сигналы имеют низкую частоту колебаний и поэтому, чтобы передать их на большое расстояние, необходимо повышать частоту информационных сигналов. Добиваются этого путем наложения информационного сигнала на другой сигнал, который имеет высокую частоту колебаний. Наложить информацию на это колебание можно путем медленного, по сравнению с периодом, изменения его амплитуды, частоты или фазы. Такой процесс называется модуляцией. В результате модуляции сигналы переносятся в область более высоких частот. Но, в свою очередь, амплитудная модуляция обладает низкой помехоустойчивостью, так как при воздействии помехи на сигнал искажается его форма, которая содержит передаваемое сообщение, что способствует снижению адекватности оценки основного тона речевого сигнала.

# Глава 3. Модификация алгоритма оценивания частоты основного тона речевого сигнала

Суть корреляционного анализа состоит в том, что автокорреляционная функция отражает периодические свойства сигнала. Для любого периодического сигнала автокорреляционная функция (4) достигает максимума в точках кратных периоду основного тона сигнала. Параллельно в качестве целевой предлагается функция минимизации суммы квадрата разности между отсчетами речевого сигнала и их предыдущими значениями.

Интерес представляет из себя первый глобальный максимум корреляционной функции, поскольку чем больше сдвиг, тем менее надежными являются следующие значения максимума за счет изменения амплитуд. Для его поиска применяется следующий подход. Пусть речевой сигнал представлен в виде последовательности отсчетов $S_i$, $i$=0,1,2... Для вокализованных сигналов можно считать, что временной вид речевого колебания почти точно повторяется на каждом очередном периоде основного тона

$$S_n \approx S_{n-T}, \tag{3}$$

где $T$ – период основного тона, выраженный в числе отсчетов.

Сделаем предположение, что энергия речевого сигнала не меняется на участке квазистационарности. Тогда оценка периода основного тона $k$ должна максимизировать корреляционную функцию

$$R(n,k) = \sum_{i=0}^{N-1} S_{n+i} S_{n-k+i}. \tag{4}$$

В момент времени $n$ выбирается значение оценки периода основного тона $k$, минимизирующее функцию, которая определяется как сумма квадратов разностей между отсчетами сигнала $(n{+}i)$ и отсчетами сигнала $(n\text{-}k{+}i)$

$$L(n,k) = \sum_{i=0}^{N-1} (S_{n+i} - S_{n-k+i})^2. \tag{5}$$

Сигнал может сохранять свою форму на данных участках, однако значение амплитуды на различных участках может значительно разли-

чаться, такова особенность речевых сигналов. В связи с этим требуется нормировка сигнала.

Модифицируем целевую функцию

$$L(n,k) = \sum_{i=0}^{N-1} (S_{n+i} - \alpha_k S_{n-k+i})^2, \tag{6}$$

где параметр $\alpha_k$ имеет смысл коэффициента усиления. Для сдвига $k$ оптимальное значение $\alpha_k$ вычисляется по формуле

$$\alpha_k = \frac{\sum_{i=0}^{N-1} S_{n+i} S_{n-k+i}}{\sum_{i=0}^{N-1} S_{n-k+i}^2}. \tag{7}$$

Существует иная модификация корреляционной функции, обладающей нормировкой по скользящему сигналу. В момент времени $n$ следует выбрать в качестве оценки периода основного тона такое значение $k$ , которое максимизирует функцию

$$M(n,k) = \frac{\left(\sum_{i=0}^{N-1} S_{n+i} S_{n-k+i}\right)^2}{\sum_{i=0}^{N-1} S_{n-k+i}^2}. \tag{8}$$

Эти методы позволяют получить достаточно точную оценку основного тона, которая плавно меняется во времени в соответствии с изменением речи.

# Глава 4. Программная реализация оценивания частоты основного тона речевого сигнала. Результат эксперимента программного комплекса.

Для нахождения основного тона с помощью модификации целевой корреляционной функции была написана программа в среде C++Builder.

Звук является оцифрованным, то есть он разбит на фрагменты, которым присвоены некоторые числовые значения. Если записывать звук через микрофон, то он переводится в электрический сигнал, напряжение которого непрерывно зависит от времени. Это напряжение называется аналоговым представлением звука. Оцифровка звука делается с помощью измерения напряжения сигнала во многих точках оси времени, перевода каждого измерения в числовую форму и записи полученных чисел в файл. Этот процесс называется сэмплированием или отбором фрагментов. Звуковая волна сэмплируется, а сэмплы (звуковые фрагменты) становятся оцифрованным звуком. Устройство сэмплирования звука называется аналого-цифровым преобразователем (АЦП или ADC, analog-to-digital converter).

Переходя к программной реализации, необходимо определиться с параметрами, входящими в задачу. Многие из них уже упомянуты выше. На вход подается файл формата WAV (Waveform Audio File Format), данный формат является стандартом для хранения аудио потока на ПК. Указывается промежуток поиска периода основного тона (минимально и максимально возможный). Для этого предварительно оценивается среднее значение периода колебаний звукового файла. Существует нюанс, что при обработке звуковых сигналов часто возникают трудности, обусловленные конечностью интервала обработки. Можно добиться улучшения качества оценки, ограничив интервал анализа, тем самым снизив влияние конечных отрезков. Таким образом, в ходе анализа речь разбивается на блоки данных. Обычно такие блоки называются окнами. Ограничение интервала анализа равносильно произведению исходного сигнала на оконную функцию. Было выбрано окно Хемминга [9], представляющее собой косинус-окно со сглаживанием на концах интервала.

Далее при разбиении сигнала на окна осуществляется корреляционный анализ, который позволяет установить в относительно небольших

оцифрованных звуковых фрагментах (далее сэмплах) наличие определенной связи изменения значений сигналов по смещению окна.

В формуле (4) имеется фрагмент $S_{n+i}$, в котором может быть (а может и не быть) некоторая последовательность $S_{n-k+i}$ конечной длины $T$, где минимальное значение целевой функции (6) является искомым. Для поиска этого значения в скользящем фрагменте $S_{n+i}$ временного окна длиной $T$ вычисляются скалярные произведения числовых значений $S_{n+i}$ и $S_{n-k+i}$, участвующие в формулах (6) и (7). Тем самым искомый фрагмент $S_{n-k+i}$ "прикладывается"к фрагменту $S_{n+i}$, скользя по аргументу $k$, и по величине скалярного произведения оценивается степень сходства сигналов в точках сравнения. Чем больше сходство (меньше), тем больше (меньше) коэффициент усиления $\alpha_k$ (7), "наказывающий"сильные отклонения и "поощряющий"малые отклонения фрагмента в скользящем рассматриваемом промежутке. Минимизируя функцию $L(n,k)$ по длине окна, взятого как двойная сумма минимального и максимального значения промежутка поиска периода основного тона, мы пробегаем по всему подаваемому входному сигналу, запоминая значение $k$, при котором целевая функция $L(n,k)$ достигла минимума, а $M(n,k)$ максимума. Полученное значение $k$ будет являться искомым периодом основного тона на данном промежутке. Далее происходит синтез звукового выходного файла в виде синусоиды с меняющимися частотой и амплитудой: амплитуда определена из соотношения квадратного корня соотношения сэмпла на период ОТ в текущем шагу, частота как обратная величина периода ОТ.

В качестве входных файлов были использованы некоторые произведения из классической музыки и элементы речевой базы PTDB-TUG [10]. Речевая база содержит 2342 предложения из корпуса TIMIT, надиктованных дикторами 10-ю мужчинами и 10-ю женщинами, были рассмотрены наиболее интересные экземпляры. Эксперимент показал, что среди разновидностей функции максимизации (минимизации) наиболее стабильным оказался вариант (6), поскольку имел меньше шумовых составляющих среди остальных.

Интересным аспектом определения периода основного тона является то, что при этом возможно ускорение речи без потери ее разборчивости. Путем удаления определенного процента периодов основного тона из каждого звука (фонемы) можно увеличить скорость генерации звуков, что анало-

гично более быстрому произнесению слов. Период основного тона звуков при этом остается постоянными. И напротив, если просто увеличить скорость воспроизведения речи, то все частоты, включая и частоту основного тона, пропорционально возрастут. Среднее ускорение приведет к очевидным искажениям, а при еще большем ускорении речь станет неразборчивой. Приборы, разработанные для моделирования быстрого словообразования, демонстрируют способность человека к восприятию речевой информации гораздо быстрее, чем обычное ее воспроизведение человеком. Ускорение речи является достоинством корреляционных методов обработки.

## 3.1. Обработка звукового сигнала

В ряде случаев возникает задача изменения частоты дискретизации сигнала, представленного в дискретном времени. Такая задача появляется в случае дискретизации параметра речевого сигнала с низкой частотой для более эффективного анализа, а для его восстановления требуется более высокая частота дискретизации. В этом случае частоту дискретизации следует повысить. Процесс повышения частоты дискретизации будет далее называться интерполяцией. Этот процесс позволит значительно снизить время выполнения алгоритма, поскольку корреляционный анализ является емким по времени за счет скалярного произведения. В программе выполняется десятикратное уменьшение сэмплов, создается новый массив входных данных, где каждый элемент является суммой каждых десяти значений числовой формы исходного фрагмента. После прохождения анализа осуществляется линейная интерполяция амплитуды, участвующей в генерации выходного сигнала.

Графическое изображение результата анализа сложного звука по составляющим его компонентам называют амплитудно-частотным спектром. На спектре амплитуду выражают в двух разных единицах: логарифмических (в децибелах) и линейных (в процентах). Если используют процентное выражение, то отсчет чаще всего ведут относительно амплитуды наиболее выраженной составляющей спектра. В этом случае ее принимают за нуль децибел, а уменьшение амплитуды остальных спектральных составляющих измеряют в отрицательных единицах.

Первая гармоника была выделена верна с небольшим количеством шумов. Мелодия узнаваема даже при наличии большого размаха часто-

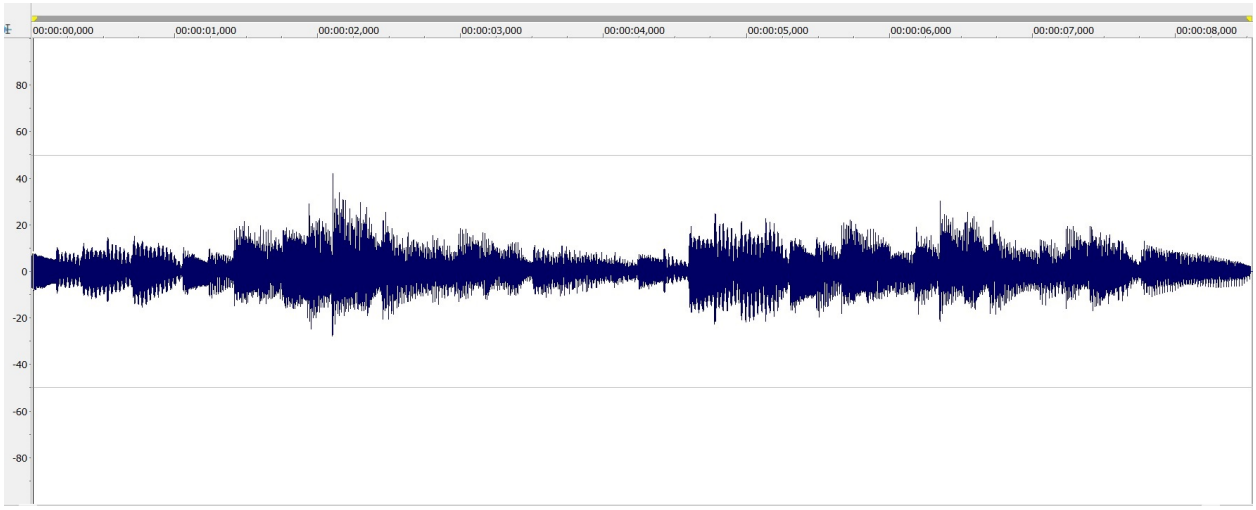


Рис. 3: Фортепианная пьеса-багатель Людвига ван Бетховена "Fur Elise"

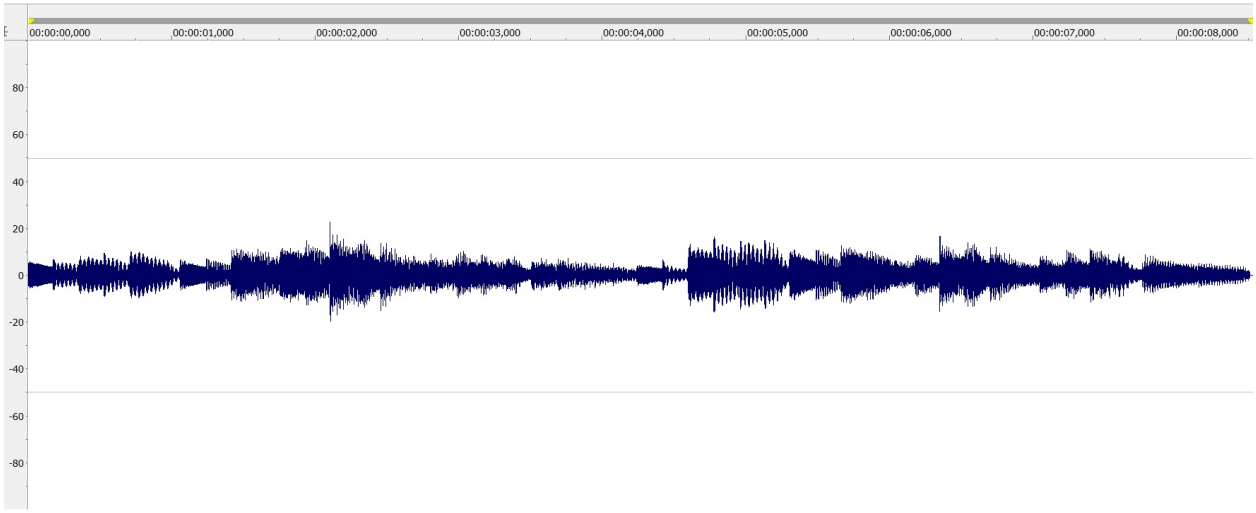


Рис. 4: Результат работы программы

ты звука от 82,407 Герц (ми большой октавы) до 659,26 Герц (ми второй октавы).

## 3.2. Обработка речевого сигнала

После успешного эксперимента с музыкальном звуком были обработаны речевые сигналы. Результаты представлены далее.

Как можно заметить, были пропущены высокие частоты, соответственно, благодаря модификации форманты не оказали сильного влияния на результат.


Рис. 5: Фортепианная пьеса-багатель Людвига ван Бетховена "Fur Elise"в приближении

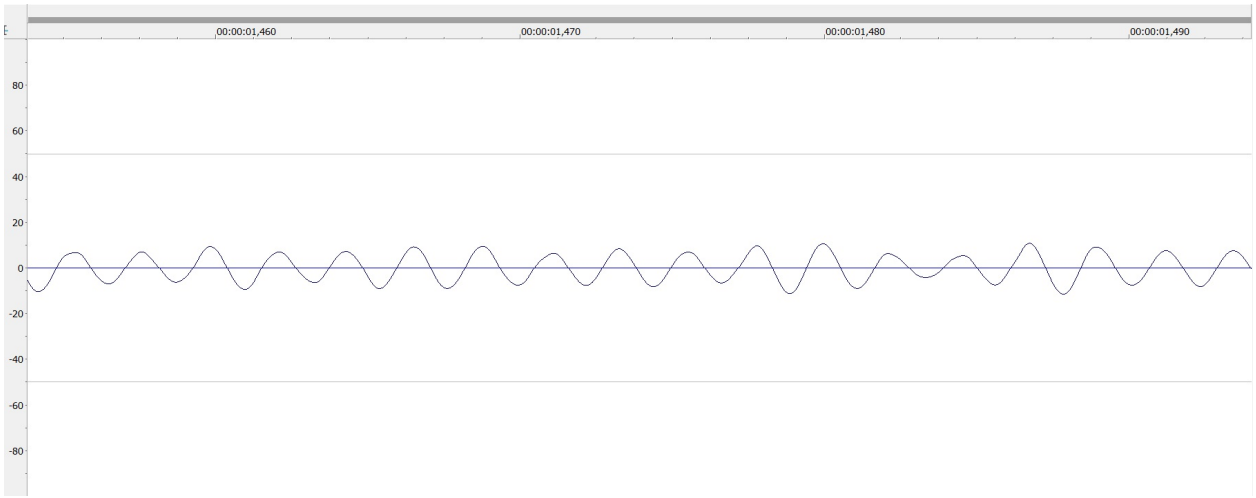


Рис. 6: Результат работы программы

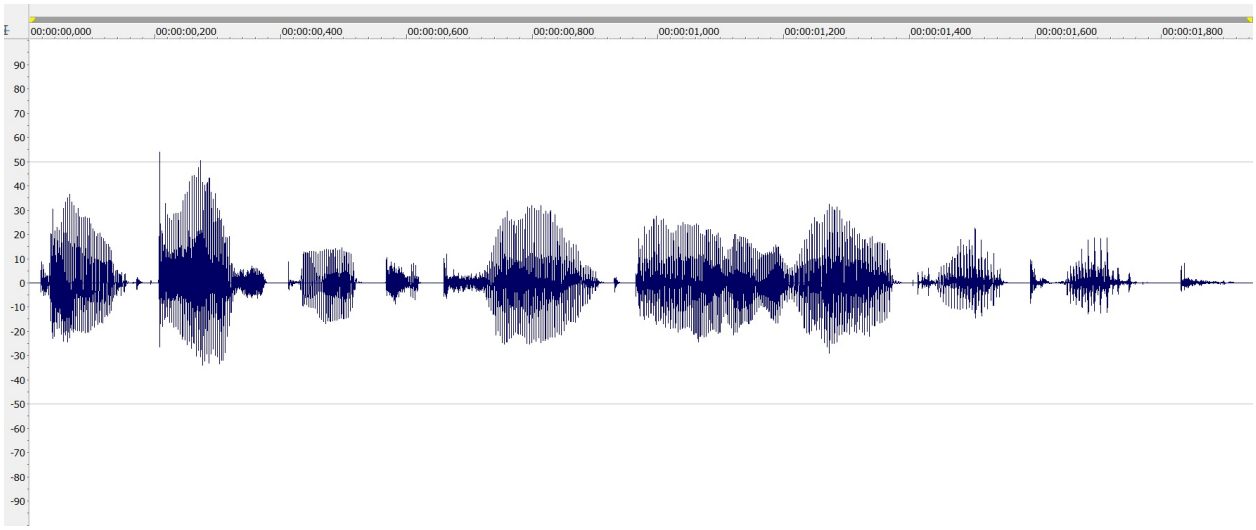


Рис. 7: Речевая фраза "Don't ask me to carry an oily rag like that"

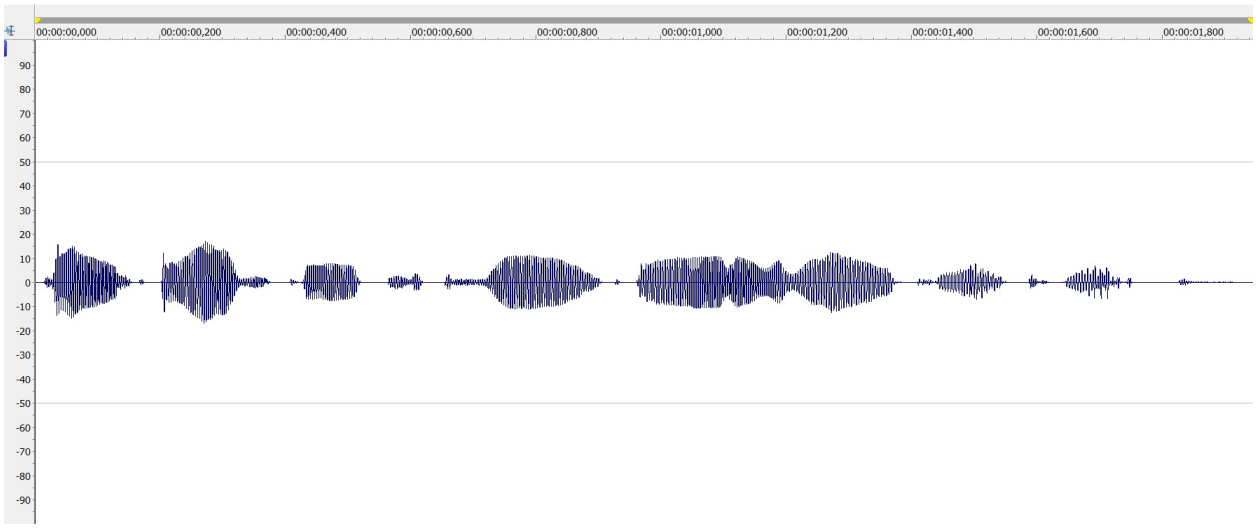


Рис. 8: Результат работы программы

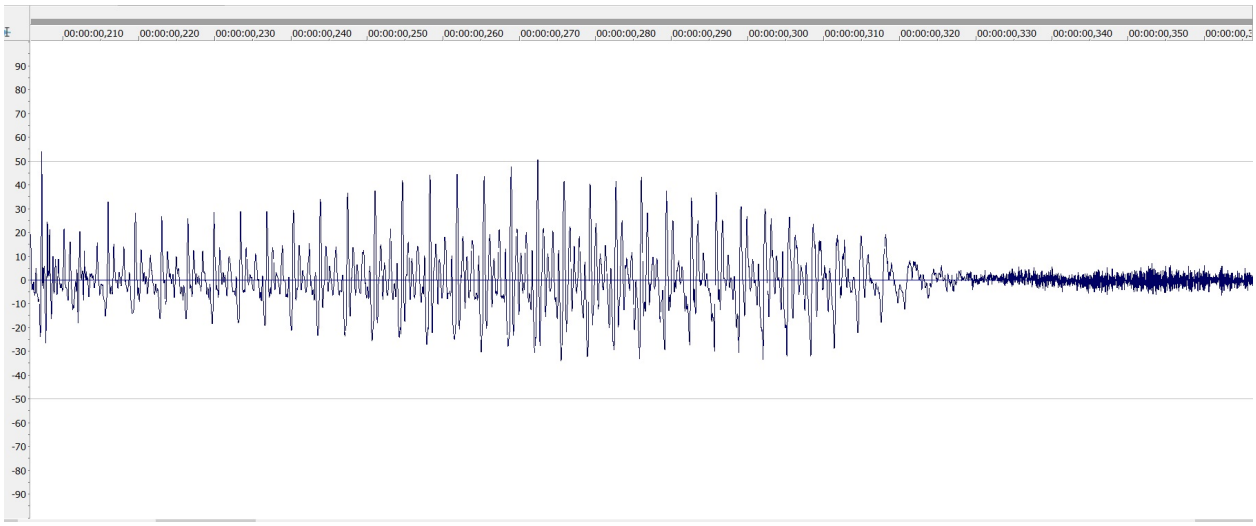


Рис. 9: Речевая фраза "Don’t ask me to carry an oily rag like that"в приближении

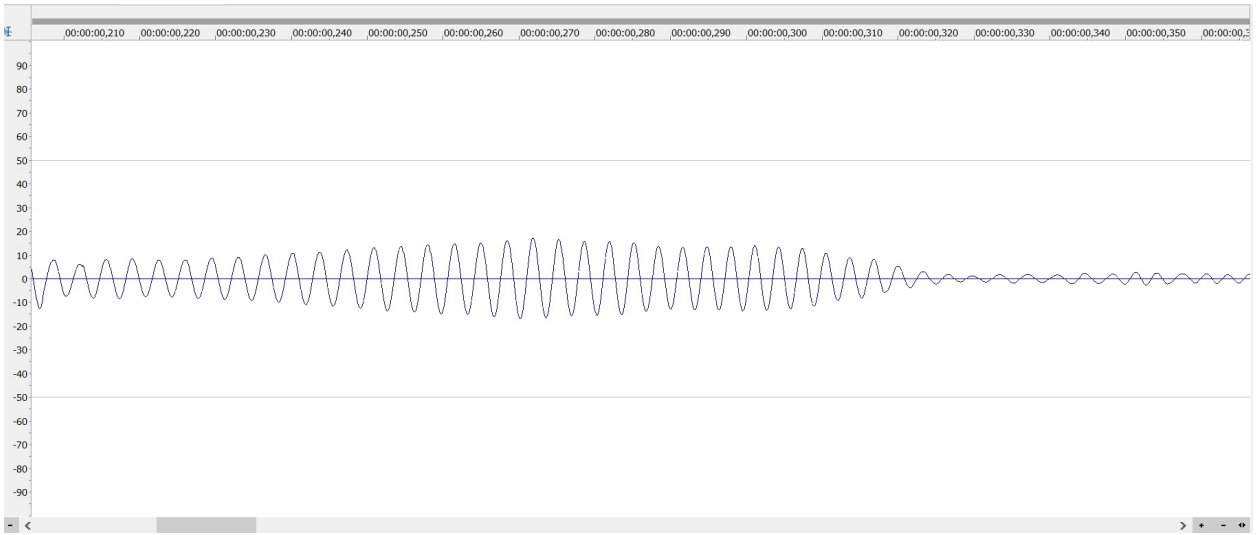


Рис. 10: Результат работы программы

# Выводы и заключение

Таким образом, в работе произведена компьютерная обработка цифрового звука с целью получения основного тона, меняющегося по времени, и определение амплитуды гармоники с основным тоном. Произведен анализ различных методов определения основного тона. Основной упор был сделан на различные виды корреляционного анализа. Исследована, в частности, вариант максимизации скалярного произведения звуковых фрагментов одинаковой длины неподвижного и скользящего. А так же, квадратичное отклонение подвижного от неподвижного. Лучшей оказалась модификация, заключающаяся в их объединении. Скалярное произведение было использовано в качестве коэффициента усиления при скользящем фрагменте. Для проверки качества обработки входного файла была создана программа акустической "иллюстрации". Она позволила быстро оценивать качество определения основного тона тем или иным способом. Наиболее показательной была проверка работы всего комплекса на простых музыкальных фрагментах, сыгранных на фортепиано.

# Список литературы


[1] Рабинер, Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер – М.: Радио и связь, 1981. – 496с.

[2] Загоруйко Н.Г. Методы распознавания и их применение. – М.: Сов. радио, 1972. – 206 с.

[3] Маркел Дж. Линейное предсказание речи / Дж. Маркел, А.Х. Грей. – М.: Связь, 1980. – 308 с.

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

[5] Гапочкин А.В. Определение основного тона речи с помощью вейвлет-преобразования и его применение//Вестник МГУП имени Ивана Федорова, 2016. – № 1. – С. 22-24.

[6] Голубинский А.Н. Расчет частоты основного тона речевого сигнала на основе полигармонической математической модели //Вестник Воронежского института МВД России, 2009. – № 1. – С. 81-90.

[7] Борискевич А.А. Электронный учебно-методический комплекс по дисциплине Цифровая обработка речи и изображений. – Минск, 2007. – 293 с.

[8] Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.

[9] Алимурадов А.К. Обзор и классификация методов обработки речевых сигналов в системах распознавания речи / А.К. Алимурадов, П.П. Чураков // Измерение. Мониторинг. Управление. Контроль. – 2015. – № 2 (12). – С. 27–35.

[10] Вольф Д.А. Модель, численная и программная реализация оценивания частоты основного тона речевого сигнала с помощью сингулярного спектрального анализа//Диссертация соискателя учёной степени к. т. н. – 2015. – С. 149

# Приложение

Реализация программного комплекса, позволяющего за приемлемое время осуществить определение основного тона в файлах типа WAV с оцифрованным звуком.

```
int indexL, indexmnk, indexS;
float L, R;
double sumL,sumR,sumMz,sumW, Mnk, minimum, maximum,
Lnk, Lbeta, beta, sumlength;
float denom, denom1, denom2, denom3, norma, nnorma, shift;
float Ampold, Amp, DeltaAmp,
deltaThetaold, DeltaTheta, deltaThetaInterp, AmpInterp;
float* W;
float* WorkTemp;
float*  WorkRun;
float* WorkMini;
int boul;
const int ten=10;
//Инициализация массивов
   fprintf(results, "SIZE=%d\r\n", SIZE);
   sumlength = 4*(minper+maxper)/ten;
   WorkMini = new float[SIZE/ten];
   WorkTemp = new float[sumlength];
   WorkRun = new float[sumlength];
   W = new float[2*sumlength];
//Окно Хемминга
   for (int i=0;i<sumlength/4;i++) //W[i]=1;
   W[i] = 0.54 - 0.46*cos(2.0*(M_PI*i)/sumlength);
   for (int i=sumlength/4;i<(sumlength*3)/4;i++)
   W[i] = 1;
   for (int i=(sumlength*3)/4;i<sumlength;i++
   W[i] = 0.54 - 0.46*cos(2.0*(M_PI*i)/sumlength);
//Заполнение сокращенного массива
       int n=0, m=0;
       for (m=0;m<SIZE/ten-10;m++)
```

```
        { WorkMini[m]=0;
         for (int i=m;i<(ten+m);i++) WorkMini[m]+=WorkR[n++];
         WorkMini[m]/=ten;
        }
       for (n=0;n<SIZE/ten-2*sumlength;n++)
       { denom=denom1=denom2=0; boul=0;  indexL=1;
            sumW=0;
            for (int i=n;i<n+sumlength;i++) sumW+= WorkMini[i];
            shift = sumW/sumlength;
            for (int i=0;i<sumlength;i++) WorkTemp[i] = WorkMini[n+i]
            - shift;
//Цикл корреляционного анализа
            for (int k = minper/ten;k <= maxper/ten;k++)
            { denom3= sumL=0; sumR=0; Lnk=Lbeta=0;
               for (int i=n+k;i<n+k+sumlength;i++) sumW+= WorkMini[i];
               shift = sumW/sumlength;
               for (int i=0;i<sumlength;i++)
                    {  WorkRun[i] = WorkMini[n+k+i] - shift;
                       R = W[i]* WorkTemp[i]* WorkRun[i]; sumR += R;
                       denom3 += W[i]* WorkRun[i]*WorkRun[i];
                     }
                              if (sumR>0)
                                 {   if (denom3<1) {
                                     fprintf(results,
                                     "denom3=%e n=%d k=%d \n",
                                      denom3, n, k);
                                     goto metka;   }
//Вычисление коэффициента усиления и целевой функции
                                   beta = sumR/denom3;
                                   if (boul==0) boul=1; else boul=2;
                                   for (int i=0;i<sumlength;i++)
                                     { Lbeta = WorkTemp[i]
                                     -beta*WorkRun[i];
                                     Lnk += Lbeta*Lbeta;
                                     }
```

```
                                    }
//Нахождение такого k, при котором было достигнуто минимальное
значение целевой функции
                                   if (boul==1)
                                    { minimum = Lnk; indexL = k; }

                                   else  if (boul==2&&sumR>0)
                                    {
                                   if (minimum>Lnk)
                                    {minimum = Lnk; indexL = k;};
                                    }
            }   //Закрытие цикла по k
//Синтез основного тона + Интерполирование
                for (int i=n;i<n+indexL;i++)
                denom+=WorkMini[i]*WorkMini[i];
                Ampold= Amp; deltaThetaold = deltaTheta;
                Amp = sqrt(denom/indexL);
                deltaTheta =  2.0*M_PI/(indexS*ten);
                DeltaAmp = Amp-Ampold;
                DeltaTheta = (deltaTheta- deltaThetaold)/ten;
                for (int j=0;j<ten;j++)
                { deltaThetaInterp =  deltaThetaold +  DeltaTheta*j;
                  AmpInterp = Ampold  +   DeltaAmp*j ;
                  thetaLocal += deltaThetaInterp;//Interp;
                  kk2 = AmpInterp * sin(thetaLocal);
                  if (n % 100 == 0)
                  thetaLocal = MinimizerSynthesis(thetaLocal);
                  SinIn[m++]=kk2;
                }
                fprintf(results,"WorkMini[n=%d]=%e boul=%d kk2=%e
                denom=%e Lnk =%e minimum = %e indexS = %d
                thetaLocal=%e   deltaTheta=%e  \r\n",
                n, (float)WorkMini[n], boul, kk2, denom, Lnk, minimum,
                indexL, thetaLocal,  deltaTheta );
        }//Закрытие цикла по n
```