

ФЕДЕРАЛЬНОЕ АГЕНСТВО ПО ОБРАЗОВАНИЮ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ

На правах рукописи

КОЦУБИНСКИЙ ВЛАДИСЛАВ ПЕТРОВИЧ

МАТЕМАТИЧЕСКИЕ МОДЕЛИ ОБРАЗОВАНИЯ ЗВУЧНОЙ РЕЧИ

Специальность 05.13.18 « Математическое моделирование, численные
методы и комплексы программ»

ДИССЕРТАЦИЯ
на соискание ученой степени
кандидата технических наук

Научный руководитель:
доктор технических наук, профессор
В.П. Бондаренко

ТОМСК 2004

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
ГЛАВА 1. МОДЕЛИ ГЕНЕРАЦИИ ВОКАЛИЗОВАННЫХ ЗВУКОВ РЕЧИ.....	16
1.1 Анатомия органов речеобразования при формировании нормальной и пищеводной речи.....	17
1.2 Акустические схемы процессов голосообразования.....	19
1.3 Диафрагма, легкие, бронхи, трахея, воздушный пузырь.....	21
1.4 Гортань, голосовые складки, псевдоголосовая щель.....	23
1.5 Модели голосообразования.....	26
1.6 Постановка задачи исследования	37
1.7 Адекватность модели.....	40
Выводы по главе:	44
ГЛАВА 2. МОДЕЛЬ ОБРАЗОВАНИЯ НОРМАЛЬНОЙ И ПИЩЕВОДНОЙ РЕЧИ.	46
2.1 Входной импеданс речеобразующего тракта.....	46
2.2 Аппроксимация входного импеданса речеобразующего тракта.....	52
2.3 Модель фонации для нормальной речи	61
2.4 Модель фонации для пищеводной речи	65
Выводы по главе:	69
ГЛАВА 3. ИССЛЕДОВАНИЕ МОДЕЛЕЙ ОБРАЗОВАНИЯ ЗВУЧНОЙ РЕЧИ	70
3.1 Оценка параметров моделей образования звучной речи	70
3.2 Параметры и роль гортанного желудочка в процессе образования звучной речи	75
3.3 Начальные условия	79
3.4 Стационарные решения	83
3.5 Исследование устойчивости процессов образования звучной речи....	91
Выводы по главе:	99

ГЛАВА 4. ИССЛЕДОВАНИЕ СТРУКТУРЫ СИГНАЛОВ ЗВУЧНОЙ РЕЧИ.....	100
4.1 Комплекс для исследования структуры речевых сигналов	100
4.2 Исследование влияния разности давлений в трахее и ротовой полости на частоту основного тона.....	107
4.3 Синтез речевых сигналов гласных звуков.....	110
4.4 Исследование спектральных характеристик	124
Выводы по главе:	133
ЗАКЛЮЧЕНИЕ	134
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	136
ПРИЛОЖЕНИЕ 1	151
ПРИЛОЖЕНИЕ 2	152

ВВЕДЕНИЕ

Исследование процессов речеобразования, а также попытки построения синтезаторов речи (говорящих машин) имеют достаточно продолжительную историю. Наиболее полно во второй половине прошлого века эти исследования были представлены в монографиях Сапожкова М.А., Фанта Г., Фланагана Дж., Сорокина В.Н. [108,116,117,125,127]. В начале работы носили в основном теоретический характер, при этом модели пытались строить на электрических аналогах цепей с сосредоточенными параметрами. С развитием цифровой вычислительной техники на основе предыдущих исследований начали интенсивно использовать численное моделирование как процесса речеобразования, так и речевосприятия. Наиболее полно эти цифровые модели и алгоритмы обработки, распознавания и синтеза речи представлены в работах [104,105].

С самого зарождения речевых исследований основное внимание уделялось компактному представлению речевого сигнала для передачи его по низкоскоростным линиям связи. Эти наработки привели к созданию ряда вокодеров, использующих в тот или иной способ сжатия речевого сигнала, описания которых приводятся в работах Мясникова Л.Л. [86].

С 60-х годов прошлого столетия начались интенсивные исследования по созданию синтезаторов и распознавателей речи [4, 11, 12 40, 50, 53, 64, 66, 82, 92, 97, 106, 107, 132]. По этой проблеме с 1965 года в СССР действовал семинар по автоматическому распознаванию слуховых образов (АРСО), а затем в РФ этот семинар получил название «Теория и практика речевых исследований», который проводится ежегодно в МГУ им. М.В. Ломоносова.

Одним из перспективных направлений является синтез речевого сигнала. Здесь можно выделить следующие направления:

- кодирование - восстановление речевого сигнала (вокодеры) [29,33];
- компилятивный синтез речи (формантный вокодер) [49,67,68,73,74];
- синтез речи по правилам (на основе моделей речеобразования) [1, 24, 33,

45, 49, 65, 67, 69, 75, 77].

Многие ученые создавали и создают синтезаторы речи. На каждом этапе развития науки и техники превалировали разные способы. Например, в 1983 году Лобановым Б.М. был создан синтезатор человеческой речи 15КС200-014 (УРМЗ.852.035), в котором на аналоговых элементах был смоделирован речеобразующий тракт. В синтезаторе использовалось несколько источников для генерации различных типов звуков. Для того времени это был первый в мире синтезатор речи, использующий синтез по правилам, который выпускался серийно. Слоговая разборчивость речевого сигнала [25, 26, 96] на выходе синтезатора удовлетворяла оценке “хорошо”. Однако речевой сигнал данного синтезатора имел низкую естественность, поэтому его разборчивость резко падала при наличии внешних акустических шумов, что существенно суживало области его применения. Это объясняется тем, что существовавшие в то время, да и в настоящее, модели только в общих чертах соответствуют реальным речеобразующим системам человека.

Поэтому уже в версиях для ЕС ЭВМ и IBM PC Фонемофон 4 и Fonemafon автор использовал компилятивный синтез. В дальнейших разработках модуль Sakrament text-to-text engine на основе Microsoft Speech API технологии Лобанов Б.М. использовал компилятивный синтез речи.

Актуальность темы диссертации. Несмотря на достигнутые успехи в синтезе (Д. Клатт 1986[149, 150], Б.М. Лобанов [73-77] 1991, 2001 – Fonemafon, Sokramento, клуб Речевых Технологий МГУ – Голосовая мышь, Агафья и Агофон), распознавании речи (Dragon Dictate и Горыныч), а также идентификации диктора по речевому сигналу, в настоящее время нерешенными остаются следующие проблемы [11,12,16,17,19]:

- не установлена связь между параметрами речеобразующей системы и характеристиками речевого сигнала;

- нет адекватных моделей речеобразования не только нормальной, но и патологически измененной (т.е. фарингеальной и пищеводной) речи;

- не решена проблема синтеза речи по правилам.

Кроме того, является актуальной задача изучения механизмов образования пищеводной речи [2, 59, 121]. Пищеводная речь является заменой нормальной речи и позволяет реабилитировать онкологических больных после полного удаления гортани, т.е. практически инвалидов вернуть к нормальной жизни.

Восстановление звучной речи у больных, перенесших ларингэктомию или резекцию гортани, находится на стыке физиологии, клинических дисциплин, специальной педагогики - так писала в 1985 году доктор биологических наук Таптапова С.Л. [121]. В книге «Восстановление звучной речи после резекции или удаления гортани» автором дана оценка эффективности комплексных методических подходов к восстановлению речи. Восстановление голоса после резекции гортани наблюдается в 88 % случаев.

Предложенные и усовершенствованные Балацкой Л.Н., Кицманюк З.Д. [2], при участии автора [59] совместно с Бондаренко В.П., в 2001 году методики восстановления звучной речи и голоса позволяют сократить сроки реабилитации и повысить эффективность голосообразующей функции. Это дает возможность вернуть к прежнему социальному статусу 67,6% пациентов после ларингэктомии, 78,2% - после резекции гортани, 86,9% после органосохранных операций органов полости и ротового отдела глотки, 98% - с парезами и параличами гортани.

Основными проблемами, с которыми сталкиваются онкологи, при восстановлении звучной речи после хирургического лечения рака гортани являются: преодоление больным психологического стресса в послеоперационном периоде, необходимость адаптации к новому типу дыхания через трахеостому, формирование нового компенсаторного фонационного органа в первом физиологическом сужении пищевода. Эта задача решается комплексно при совместной работе онколога, логопеда, психотерапевта или психолога и инструктора по лечебной физкультуре.

Для интенсификации процесса обучения больных пищеводной речи необходимо определить, какого объема и где образуется воздушный пузырь, то

есть целенаправленно определять методику обучения. Частично эти вопросы можно решить путем математического моделирования процессов в речеобразующей системе человека при образовании нормальной и пищеводной речи.

В настоящее время все созданные модели речеобразования базируются на результатах научных исследований, проведенных более 50 лет назад Г.Фантом, миоэластической теории, сформулированной Ferrein в 1741 г, нейроронаксической теории обоснованной Husson в 1950 г. Заслуживает внимания и мукоондуляторная теория фонации, выдвинутая J. Perello (1962). Основопологающим в названных теориях является положение о независимости источника возбуждения (голосовых складок) от процессов, происходящих в голосообразующей системе. Эти положения привели к появлению формантной теории голосообразования, которая была использована при создании формантных вокодеров и синтезаторов речи по правилам. Огромное количество экспериментальных исследований, проведенных в этом направлении, показало, что речевой сигнал на входе формантных вокодеров и формантных синтезаторов голоса резко отличается от естественного сигнала. Все это, в частности, привело к тому, что в мультимедийных системах синтеза голоса используется метод компиляции.

Использование этих теорий ведет к тому, что исследователи не достаточно уделяли внимание влиянию некоторых органов речеобразующей системы на процесс образования звучной речи. Поэтому для создания адекватной модели работы системы синтеза речи следует отказаться от каких-либо ограничений и в тоже время избегать чрезмерного внимания к каждому органу в отдельности. Данная работа основана на исследованиях (1973) Галунова В.И. [30], который впервые экспериментально показал, что речевой сигнал сформирован уже на выходе голосовой щели. Вышедшая в 2002 году книга Морозова В.П. [84], в которой обсуждается необходимость рассмотрения голосообразующей системы целиком (не только от голосовых складок до губ), объясняет важность исследований в данном направлении. При этом он для

обоснования своего подхода использовал и результаты исследований, полученные автором [55-57].

Таким образом, данная работа посвящена решению актуальной проблемы, а именно созданию моделей, использование которых для синтеза речевого сигнала приведет к генерации всего многообразия звучной речи и позволит более эффективно подходить к реабилитации больных после резекции гортани. При этом основное внимание уделяется моделям образования звучной речи, то есть той речи, которая формируется с участием голоса при работе голосовых складок. Необходимо учитывать, что более 60-70% (по времени) речевого сигнала формируется с участием голоса, эти части определяют естественность и узнаваемость речи, и поэтому очень важно, чтобы реабилитируемые больные овладели звучной речью. Это определяет актуальность диссертационной работы и ее соответствие современным тенденциям развития перспективных систем синтеза речи по правилам.

Цель работы и задачи исследования. Основной целью диссертационной работы является построение и исследование моделей образования нормальной и пищеводной звучной речи, выявление механизмов генерации речевых сигналов с учетом работы голосовых складок. Для достижения поставленной цели в диссертационной работе были сформулированы следующие задачи:

- 1) построение электроакустических аналогов и соответствующих им моделей образования нормальной и пищеводной речи;
- 2) анализ моделей с целью установления связи между электрическими, механическими и акустическими параметрами речеобразующей системы;
- 3) исследование процессов, возникающих в моделях речеобразования, и определение условий устойчивой генерации речевого сигнала при образовании нормальной и пищеводной речи;
- 4) генерация речевых сигналов вокализованных звуков речи и сравнение их реальными речевыми сигналами.

Методы исследования. Для решения поставленных задач в работе используются математический анализ, теория дифференциальных уравнений,

теория электрических цепей, теория управления, методы теории восприятия и разборчивости речи, методы спектрального, синхронного с частотой основного тона анализа речи.

Научная новизна. Научная новизна работы состоит в следующем:

- 1) математические модели образования нормальной и пищеводной речи, позволяющие исследовать процессы генерации вокализованных звуков;
- 2) результаты исследования процессов образования нормальной и пищеводной звуочной речи, а также границы возникновения устойчивых колебаний в речеобразующей системе при генерации гласных звуков;
- 3) алгоритмы синтеза речевых сигналов гласных звуков и результаты сопоставительного анализа сгенерированных в моделях речевых сигналов с реальным.

Тезисы, выносимые на защиту.

- 1) математические модели образования нормальной и пищеводной речи, рассматриваемые как системы в целом: легкие, бронхи, трахея, голосовые складки, речеобразующий тракт;
- 2) математические модели устойчивых колебаний движения голосовых складок в полной речеобразующей системе;
- 3) математические модели генерации речевых сигналов вокализованных звуков.

Практическая ценность работы. В данной работе были разработаны следующие алгоритмы и модели:

- 1) алгоритмы синтеза по правилам естественных речевых сигналов гласных звуков;
- 2) акустические модели голосообразования нормальной и пищеводной речи, позволяющие целенаправленно проводить обучение восстановленного голоса после резекции гортани;
- 3) построены и исследованы теоретические модели голосообразования вокализованной нормальной, пищеводной и фарингеальной речи;
- 4) алгоритмы численного моделирования процесса образования нормальной и пищеводной речи;

- 5) разработаны методы оценки параметров анатомии, физиологии и нейродинамики процессов голосообразования и сопоставление их с результатами моделирования;
- 6) уточнены теоретические модели с учетом результатов клинических исследований;
- 7) изучение особенностей поведения системы голосообразования (возникновение устойчивых колебаний голосовых складок или их эквивалентов, возникновение хаотичных колебаний, особых точек пространства состояния).

Полученные результаты позволяют ставить и решать задачи по синтезу естественной речи по правилам, обоснованно подходить к разработке алгоритмов описания речевых сигналов, устанавливать зависимости между анатомией диктора и его голоса.

Реализация результатов работы. Алгоритмы и отдельные методики, разработанные в диссертационной работе, внедрены в НИИ онкологии Томского научного центра СО РАМН, в учебный процесс на факультете вычислительных систем на кафедре, КСУП и на факультете электронной техники на кафедре, ПрЭ Томского государственного университета систем управления и радиоэлектроники.

Материалы 2 и 3 глав диссертации составили основу пособия для студентов вуза по курсу «Передача данных в информационно-управляющих системах».

Личный вклад автора. Разработаны акустические схемы образования нормальной пищеводной и фарингеальной речи, совместно с научным руководителем д.т.н., проф. Бондаренко В.П., сформулирована постановка задачи и разработаны модели образования нормальной и пищеводной речи.

Лично автором проведен анализ устойчивости, а также исследование пищеводной речи. Разработано техническое задание на комплекс для исследования структуры речевых сигналов, который был разработан Пономаревым А.А.

Проведено исследование структур сигналов нормальной и пищеводной речи.

Апробация результатов работы. Основные положения и результаты диссертационной работы представлялись на международных семинарах «Речь и Компьютер» SPECOM (Санкт-Петербург, 2000; Москва, 2001; Санкт-Петербург, 2002), на международных научно-практических конференциях «Природные и интеллектуальные ресурсы Сибири» (Сибресурс) (Новосибирск, 1996; Красноярск, 1997; Барнаул, 1998; Омск, 1999; Тюмень, 2000; Барнаул, 2001), на конференциях Теория и практика речевых исследований (АРСО), проходивших в МГУ им. М.В. Ломоносова (1999 и 2001), на XI и XIII сессиях акустического общества (Москва, 2001 и 2003), на Нижегородской акустической научной сессии (Нижний Новгород, 2002), на конференциях «Управляющие и вычислительные системы. Новые технологии» Вологда, 2000) и «Системные проблемы качества, математическое моделирование информационных, электронных и лазерных технологий» (Сочи, 2002).

Большая часть работы опубликована в сборниках научных статей «Методы и алгоритмы автоматизации технологически процессов» (Томск: Изд-во ТГУ 1995), «Автоматизация проектирования, идентификация и управление в сложных системах» (Томск: Изд-во НТЛ 1997), «Автоматическое и автоматизированное управление сложными системами» (Томск: Изд-во ТГУ 1998), «Интеллектуальные автоматизированные системы проектирования, управления и обучения» (Томск: Изд-во ТГУ 2000) под ред. проф., д.т.н. Тарасенко В.П. Также статьи были опубликованы в сборниках «Интеллектуальные системы в управлении, конструировании и образовании» издаваемых (Томск: Изд-во STT 2000-2004) под ред. проф., д.т.н. Шелупанова А.А., «Измерения, автоматизация и моделирование в промышленности и научных исследованиях» (Бийск: мзд-во АлтГТУ, 2001 и 2002) под ред. проф., д.т.н. Г.В.Леонова.

Публикации. Основные результаты по материалам диссертационной работы опубликованы в 34 печатных работах. Из которых 18 докладов и 16

статей.

Структура и объем работы. Диссертация, объемом 156 машинописных страницы, содержит введение, четыре главы, заключение, список литературы (173 наименования), 22 таблицы, 97 рисунков, два приложения.

Содержание работы

Во введении обоснована важность и актуальность темы диссертации, сформулированы цель работы и основные задачи, которые необходимо решить для ее достижения, характеризуется научная новизна и практическая ценность работы, кратко излагаются основные результаты работы.

В первой главе проводится краткий анализ анатомии и физиологии системы речеобразования и приводится сопоставление системы речеобразования нормальной и пищеводной речи.

Отмечается, что особенностью акустической схемы фонации нормальной речи является наличие особого образования сразу после складок – гортанного желудочка (морганьевого желудочка).

Развернутая постановка задачи концентрируется относительно двух гипотез:

1. Нейрохронокстической гипотезы, в которой движение складок объясняется нейронной активностью. В самой простой модели процесс фонации управляется нервными импульсами, т.е. движение голосовых складок не зависит от разности давлений в трахее и речеобразующем тракте.

2. Миоэластической теории, которая объясняет колебательные движение голосовых складок на этапе раскрытия подскладочным давлением, а на этапе схлопывания - силами Бернулли.

При постановке задачи обращается внимание, что на колебательные движения голосовых складок может оказывать влияние как подскладочное, так и надскладочное давление. В этом случае существенное влияние на движение голосовых складок будут оказывать выходной импеданс трахеи и входной импеданс речеобразующего тракта $Z_{\text{рот}}$. В такой постановке колебательные движения голосовых складок приводят к изменению подскладочного и

надскладочного давлений, которые в свою очередь влияют на их движения. Тогда существенной задачей является анализ устойчивости колебаний в речеобразующей системе в целом.

Во второй главе формируются электроакустические аналоги речеобразующего тракта.

Сорокин В.Н. [117] указывал следующее: «При построении электрических аналогов рассмотренных акустических схем необходимо учитывать соотношения между размерами речеобразующих систем l и длиной волны λ колебательных процессов в них. Если $l < \lambda$, то система имеет характер системы с сосредоточенными параметрами, если $l > \lambda$ реально, много больше, то система приобретает свойства системы с распределенными параметрами. В частности, при максимальной длине голосообразующего тракта 0,2 м и скорости звука 350 м/с имеем, что для частот до 1750 Гц $l < \lambda$ и только для частот больше 1750 Гц $l > \lambda$. Практически всегда процессы генерации импульсов основного тона, т.е. процесс фонации, а также формирование формант можно рассматривать на основе электрических аналогов с сосредоточенными параметрами».

Показывается, что электрические модели речеобразующего тракта, основанные на цепях с распределенными параметрами и с сосредоточенными параметрами, в диапазоне частот до 5 кГц эквивалентны. Это позволило моделировать речеобразующий тракт цепной схемой из пяти Т-образных звеньев. На основе данных по формантной структуре речевого сигнала, с учетом известных сечений речеобразующего тракта (согласно Сорокину В.Н. [116]), были определены параметры этих цепей.

По результатам анализа, проведенного в первой главе, было показано, что при аппроксимации коэффициентов передачи речеобразующего тракта, в первом приближении, можно ограничиться учетом частот первых пяти формант. Тем более что достоверных данных по частотам более высоких формант практически нет, а уровень речевого сигнала для более высоких частот очень низок.

Анализ резонансных свойств модели входного импеданса речеобразующего тракта показывает, что он приближается по своим характеристикам к реактивному двухполоснику. В этом случае задача аппроксимации заданного импеданса сводится к синтезу соответствующего реактивного двухполосника.

Показано, что в качестве модели речеобразующего тракта может использоваться электрический аналог на сосредоточенных параметрах, это позволяет свести задачу к решению системы обыкновенных дифференциальных уравнений.

Предложен подход к оценке параметров электрического аналога речеобразующего тракта. Выбрана топология электрического аналога схемы замещения схемы, и проведена оценка значений элементов этой схемы.

Построены модели голосообразования для нормальной и пищеводной речи. Выявлены качественные различия в процессах голосообразования нормальной и пищеводной звуковой речи. Показано, что следует ожидать существенного повышения уровня высокочастотных компонентов в пищеводной речи, а также существенной неустойчивости частоты основного тона.

В третьей главе производится оценка параметров речеобразующей системы и устанавливаются соотношения между акустическими и электрическими параметрами, определяются их размерности.

Выявляется роль объема гортанного желудочка, показывается, что он может оказывать существенное влияние на движение голосовых складок. Демонстрируется, что в нем возникает дополнительный резонанс, который приводит к стабилизации колебаний голосовых складок. Это косвенно подтверждается известными данными о подвижности гортани при фонации.

Проводится анализ начальных условий и поведения системы речеобразования в стационарном режиме. Показывается, что при определенных условиях возможно возникновение воздушного потока в обратном направлении. Это позволило определить граничные значения между потерями в

трахеи, речеобразующем тракте и сопротивлением голосовой щели. Проводится оценка параметров модели речеобразования с учетом известных данных по площади раскрытия голосовой щели, устанавливаются связи между этими параметрами.

Анализ устойчивости проводился численными методами в среде MathCAD. Определялись собственные значения соответствующих матриц, и анализировались их действительные части. При этом считалось, что если все действительные части собственных значений отрицательны, то система имеет устойчивые колебательные решения [70,71].

В результате численного моделирования определены области возникновения устойчивых колебаний для нормальной и пищеводной речи в зависимости от параметров легких, трахеи, гортанного желудочка, речеобразующего тракта, воздушного пузыря и псевдоголосовой щели.

Четвертая глава посвящена результатам моделирования процессов в речеобразующем тракте в виде диаграмм токов и напряжения. В ней кратко описан комплекс для исследования структуры речевого сигнала.

Проведено исследование структуры речевых сигналов, в частности его основной характеристики – частоты основного тона.

Измерения частоты основного тона на слогах реальных речевых сигналов косвенно подтверждают положение, взятое за основу моделей речеобразования, о том, что колебания голосовой щели определяются разностью давлений в трахее и ротовой полости.

Исследования структуры различных речевых сигналов, как реальных так и синтезированных, позволяют утверждать что предложенные модели адекватны, в первом приближении, и соответствуют реальным механизмам речеобразования. Сравнение синтезированных и реальных речевых сигналов позволяет сделать вывод, что исследованные модели речеобразования адекватны реальным системам. Это подтверждается субъективным восприятием синтезированных сигналов, а также качественным их динамическими спектрограмм.

ГЛАВА 1. МОДЕЛИ ГЕНЕРАЦИИ ВОКАЛИЗОВАННЫХ ЗВУКОВ РЕЧИ

Под голосообразующей системой человека следует понимать целый комплекс органов, которые принимают участие в процессе генерации речевого сигнала. К таким органам относятся: диафрагма, легкие, бронхи, трахея, гортань и глотка со всеми ее разделами, носовая полость и ее придаточные пазухи, органы полости рта.

В моделях [30,67,68,69,76,77,108,114,117,125,127,149] предполагается, что эта система может быть представлена в виде двух относительно независимых подсистем, границей между которыми являются голосовые складки, а именно:

- а) подскладочная, включающая легкие, бронхи и трахею;
- б) надскладочная, включающая гортань, ротовую и носовую полости, которые будем называть речеобразующим трактом.

Эта граница кажется естественной, и в моделях речеобразования полагается, что в первой подсистеме формируется избыточное давление, характер которого определяет мелодику и ритмику речи, а во второй формируется спектральный состав звуков. Поэтому функционирование этих подсистем рассматривается независимо [76,77,108,117,125,127], или же их взаимодействие учитывается введением поправок, связанных с работой голосовых складок [115,116,117,126,127,149] в основном в фазе раскрытой голосовой щели.

Правомерность такого подхода к моделированию нельзя считать полностью обоснованной, так как управление легкими, бронхами, трахеей, речеобразующим трактом в процессе речеобразования достаточно сложно и более взаимосвязано, чем предполагается в моделях. Вероятно, что ряд ответов на эти вопросы можно получить на основе сопоставительного анализа естественной и пищеводной речи. Голосообразующая система для пищеводной речи в результате хирургического вмешательства изменяется (удаляется

гортань), а речеобразующий тракт остается без изменений. Это позволит исследовать влияние обеих подсистем на процесс голосообразования.

Естественно, что предварительно надо провести анализ особенностей анатомии естественной и пищеводной голосообразующих систем. Это позволит более целенаправленно проводить сопоставления речевых сигналов (особенно вокализованных), полученных при разных способах речеобразования.

1.1 Анатомия органов речеобразования при формировании нормальной и пищеводной речи

Органы, участвующие в процессе речеобразования, представлены на рис. 1.1. [38,121].

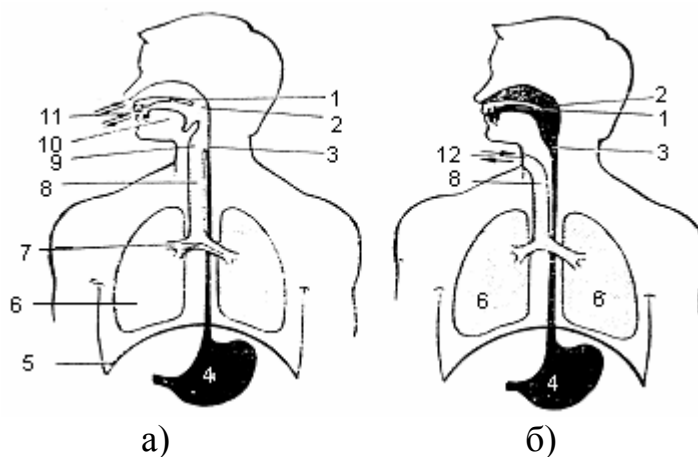


Рисунок 1.1. Схема расположения органов голосообразования до (а) и после (б) ларингэктомии (1 – небо, 2 – фарингкс, 3 – пищевод, 4 – желудок, 5 – диафрагма, 6 – легкие, 7 – бронхи, 8 – трахея, 9 – гортань, 10 – язык, 11 – губы, 12 – трахеостом)

Для нормальной речи основой голосообразования является избыточное давление в легких, которое образуется за счет уменьшения объема после вдоха. Это давление через бронхи и трахею передается на голосовые складки. Таким образом, подскладочная система ответственна в первую очередь за создание избыточного давления, обеспечивающего процесс голосообразования необходимой энергией. Для нормального развития данных процессов требуется скоординированная работа всех мышц, управляющих гортанью и голосовыми

складками, и все это должно быть скоординировано с артикуляцией речеобразующего тракта.

Образование пищевого голоса (рис. 1.1б) представляет собой сложный комплексный процесс, создания компенсаторного фонационного органа, при сохранности артикуляционного органа и нейрофизиологических механизмов речи акт, в котором ведущую роль играет нервная деятельность [38]. Речь идет о последовательности действий у больных, перенесших ларингэктомию, для фонации. Во-первых, момент заглатывания воздуха в пищевод, во-вторых, образование на уровне V – VI шейных позвонков [2,37,38] псевдоголосовой щели в I физиологическом сужении пищевода и выработке ее произвольного смыкания, что служит основой появления пищевого голоса. В пищеводе образуется воздушный пузырь, в котором создается избыточное давление, являющееся эквивалентом избыточного давления в легких при образовании нормальной речи. Сложность процесса определяется тем, что его создание требует выработки новых безусловно-условных рефлекторных связей и происходит в условиях затрудненной перестройки функции внешнего дыхания – через трахеостому и воздействия на больных ряда внешних травмирующих



Рисунок 1.2 Рентгенограмма больного С., 41 год, Фонация пищевым голосом[2]



Рисунок 1.3 Рентгенограмма больного К., 41 год, не овладевшего звуочной речью [2]

факторов соматогенного и психогенного характера. В работе [121] дается классификация видов фонации после удаления гортани:

- фонация при выработке нового физиологического образования на уровне V и VI шейных позвонков – псевдоголосовой щели, т.е. пищеводный голос (рис. 1.2).
- фонация без какого-либо нового физиологического образования только при наличии утрированной артикуляции и сближения мягкого неба с задней стенкой глотки при резком подъеме корня языка, т.е. фарингеальный голос (рис. 1.3).

Эти рисунки предоставлены ведущим научным сотрудником отделения опухолей головы и шеи НИИ Онкологии Томского научного центра СО РАМН д.б.н. Балацкой Л.Н. На рис. 1.2 можно видеть сужение пищевода на уровне C₄₋₅ под ним находится воздушный пузырь.

Таким образом, артикуляционные органы при фарингеальной используются для утрированной артикуляции [2,37,38], а в случае нормального голосообразования глотка используется для завершения формирования голоса. Проведенный краткий анализ анатомии органов речеобразования позволяет составить соответствующие акустические схемы, на основе которых проведем предварительный анализ различий в процессе голосообразования нормальной и пищеводной речи.

1.2 Акустические схемы процессов голосообразования

Анализ рис. 1.1, 1.2 и 1.3, а также литературных источников [2,37,38,51,77, 84, 85, 86, 87, 90, 108, 117, 125, 127, 130] позволяет свести возможные акустические схемы фонации к совокупности взаимодействующих объемов и отрезков труб постоянного или переменного сечения [35, 48, 52, 72, 78, 91,102]. Акустические схемы при формировании нормальной, пищеводной и фарингеальной речи сведены в таблицу 1.1, где использованы следующие обозначения:

V – объем легких, $V_{\text{п}}$ – объем воздуха в пищеводе, $V_{\text{вп}}$ – объем воздушного пузыря, $l_{\text{т}}$ – обобщенная длина трахеи, $l_{\text{рот}}$ – обобщенная длина речеобразующего тракта (РОТ), $l_{\text{м}}$ – длина гортанного (морганьевого) желудка, $l_{\text{вп}}$ – длина воздушного пузыря, l_1 – длина не подвижной части пищевода, $l_{\text{рп}}$ – длина ротовой полости, $h_{\text{гщ}}$ – высота голосовой щели, h – высота излучателя, P_0 – подсвязочное давление, P – надсвязочное давление.

Таблица 1.1

Функция	Форм-ние подсвязочно го давления	Форм-ние импульсов возбуждения	Механизм фонетического оформления	Акустические аналоги речеобразующей системы
Нормальная речь	Легкие с трахеей	Голосовые складки	Речеобразующий тракт	
Пищеводная речь	Формирование ниже псевдоголосовой щели воздушного пузыря [121]	Псевдоголосовая щель	Верхний отдел пищевода	
Фарингеальная речь	Нет. Используется воздух из ротовой полости	Сближение мягкого неба с задней стенкой глотки при резком подъеме корня языка	Утрированная артикуляция	

Из табл. 1.1 следует, что особенностью акустической схемы фонации нормальной речи является наличие особого образования сразу после голосовых складок – гортанного желудочка. Имеются все основания полагать, что этот объем играет существенную роль в процессе фонации [2,37,38,121]. Так, в [38] приведены томографии гортани при фонации, из которых следует, что

проявляется нарушение симметричности томографической картины отдельных участков голосового аппарата во время фонации (истинных и ложных голосовых складок, гортанного желудочка) в виде изменений в областях надскладочного и подскладочного пространства. Здесь следует иметь в виду то, что в норме при формировании низких звуков (например, гласный – «У») полость гортанного желудочка расширяется, при формировании высоких звуков (например, гласный «И») она сжимается.

В дальнейшем будем считать, что речеобразующий тракт, то есть глотка, начинается не от голосовых складок, как обычно считается при моделировании процессов речеобразования [30,67,68,69,77,108,117,125,127,149], а сразу после гортанного желудочка. Это обусловлено тем, что объем гортанного желудочка непосредственно примыкает к голосовым складкам и отделяется от остальной части ротоглотки достаточно резким изменением сечения речеобразующего тракта. Кроме того, в этом случае естественно выделяется общая часть для систем речеобразования нормального и пищеводного голоса. Эти общие части на акустических схемах табл. 1.1 выделены штриховкой.

Из приведенных акустических схем следует, что фарингеальный голос формируется в сильно искаженной системе речеобразования. Поэтому в дальнейшем ограничимся построением моделей для нормальной и пищеводной речи.

Анатомия и физиология органов, соответствующих системе речеобразования, достаточно полно представлены в различных монографиях [2,38,90,108,121,125,127,130]. Поэтому достаточно обратить внимание только на отличающиеся части систем голосообразования и особенности их общих частей. Модель речеобразующего тракта детально будет рассмотрена во второй главе.

1.3 Диафрагма, легкие, бронхи, трахея, воздушный пузырь

Легкие имеют форму конуса, верхушка которого обращена кверху, а основание располагается на диафрагме. Легкие выполняют как, дыхательную

так и голосообразующую функцию. Расширение грудной клетки, достигаемое сокращением дыхательных межреберных мышц, и опущение и уплощение купола диафрагмы, достигаемые сокращением мышечной части диафрагмы, влекут за собою расширение легких и заполнение их воздухом. Сужение грудной клетки при помощи сокращения выдыхательных межреберных мышц и переход диафрагмы в расслабленное состояние, при котором ее купол заметно поднимается, достигаются сокращением мышц брюшного пресса. Это обеспечивает сокращение гладкой мускулатуры и эластической ткани легких и уменьшает объем легочной ткани. Таким образом, осуществляется выход воздуха из легких. Средняя сила вдоха 80-150 мм рт.ст., а сила выдоха 50-60 мм рт.ст.

Трахея — цилиндрическая трубка длиной 12—17 см, состоящая из 16—20 гиалиновых хрящевых колец подковообразной формы. Диаметр трахеальных колец равен 2—3 см. Физиологические особенности позволяют трахее менять свою форму и, как следствие, длину и объем [51,117,127].

На уровне IV и V грудных позвонков трахея делится на два главных, или основных, бронха. Все разветвления бронхов, начиная от главного и кончая конечными бронхиолами, входят в состав так называемого бронхиального дерева [38,117]. Бронхиальное дерево служит проведению вдыхаемого и выдыхаемого воздуха, а альвеолярное дерево служит целям газообмена и формирования совместно с легкими турбулентного источника. Сводные данные по размерам трахеи и бронхов представлены в табл. 1.2.

Таблица 1.2 Геометрические размеры голосообразующей системы

Орган	Средняя длина, см		Площадь сечения, S_i , см ²	
	Min	Max	Min	max
Трахея	12	17	2	3
Левый бронх	6	8	1	4
Правый бронх	3	6	1,8	7

Объем легких лежит в пределах 1800 – 7200 см³, при этом порядка 10–15% объема используется для речеобразования. Достоверных данных по объему воздушного пузыря практически нет. Некоторые оценки можно сделать на

основе рис. 1.2 и 1.3. Объем воздушного пузыря составляет около $150\text{--}200\text{ см}^3$, (если сопоставить с объемом речеобразующего тракта, то есть на один два порядка меньше объема легких) и составляет 150 мл. согласно [88]. Кроме того, отрезок пищевода между воздушным пузырем и речеобразующим трактом много короче длины трахеи. Все это, по-видимому, существенно влияет на процесс фонации при образовании пищеводной речи, а именно нарушает стабильность колебаний частоты основного тона.

1.4 Гортань, голосовые складки, псевдоголосовая щель

Гортань представляет собой сложное образование, состоящее из хрящей, мышц, складок и нервного аппарата. Схема ее строения детально рассмотрена в работах [37,38,117,127,130]. Гортань подвижно соединена с подъязычной костью и подвешена вместе с ней к нижней челюсти и основанию черепа. От положения подъязычной кости, нижней челюсти, головы и корпуса зависит положение гортани. Это имеет большое значение при выполнении ею сложных функций. К подъязычной кости прикреплен корень языка, и благодаря этому гортань связана с движениями языка, главным образом в переднезаднем направлении. Кроме того, она прикреплена к грудной кости, что делает ее очень мобильным органом. Гортань при спокойном состоянии занимает определенный уровень от IV до VI шейных позвонков. У детей она расположена выше (верхний край ее находится у III шейного позвонка), у стариков опущена до уровня VII шейного позвонка.

Положение гортани не является постоянным, при глотании она значительно поднимается, при вдохе слегка опускается, при выдохе приподнимается. В работе [117] отмечается, что при фонации гортань выше, чем при дыхании, и положение ее зависит от произносимого звука. Позади гортани находится позадигортанный отдел глотки — гипофаринкс, а начиная с уровня перстневидного хряща (VI шейный позвонок) глотка переходит в пищевод. Размеры гортани зависят от пола, возраста и индивидуальных

особенностей человека. У мужчин она примерно на $\frac{1}{3}$ больше, чем у женщин; хрящи женской гортани имеют меньшую толщину.

Собственно гортанные мышцы (внутренние) делятся на суживатели и расширители голосовой щели. К последним относится перстнечерпаловидная задняя мышца. Эта мышца является основным расширителем и, таким образом, осуществляет дыхательную функцию. К расширителям голосовой щели относятся также мышечные пучки поперечной мышцы и боковые перстнечерпаловидные мышцы. Сложный механизм голосообразования обеспечивается преобладающим большинством мышц суживателей. В закрывании голосовой щели главным образом принимают участие две мышцы: перстнечерпаловидная и поперечная межнечерпаловидная и отчасти парная косая [38,117]. Натяжение голосовых складок осуществляет парная перстневидная мышца, при сокращении которой щитовидный хрящ наклоняется кпереди. Она принимает большое участие в образовании фальцетного регистра. Собственно голосовой мышцей является внутренняя часть щиточерпаловидной, она заложена в голосовой складке [38]. Строение голосовой мышцы сложно — ее волокна идут в различных направлениях: в горизонтальном, вертикальном и косом. Кроме того, часть этих волокон начинается и заканчивается в самой мышце. Такое строение обеспечивает сокращение отдельных участков голосовой складки, а также изменение ее формы, упругости и напряжения в целом, что имеет большое значение в процессе голосообразования при воспроизведении различных динамических оттенков (филировки и др.). Следует отметить, что совершенно изолированного действия мышц в действительности не бывает. Только при совместном комбинированном действии суживателей и расширителей голосовой щели и мышц, фиксирующих хрящи гортани, могут быть выполнены различные функции гортани, в частности самая сложная из них — фонаторная. Детально взаимодействие мышц гортани описано в работах [38,117].

Иннервируется гортань двумя нервами — верхнегортанным, который является в основном чувствительным, и нижнегортанным, заключающим в себе

главным образом двигательные волокна почти для всех мышц гортани. Гормональные нарушения в соответствующие возрастные периоды служат причиной изменения фоноторных функций гортани. Изменения, возникающие во многих эндокринных железах, в частности щитовидной, надпочечных, половых и гипофизах, являются физиологическими возбудителями гортанных мышц.

Дыхательная функция гортани осуществляется следующим образом. При вдохе голосовые складки расходятся, образуя щель в виде равнобедренного треугольника, которая может изменяться в зависимости от интенсивности дыхания, т. е. силы вдоха и выдоха. При форсированном вдохе голосовые складки максимально расширяются и иногда, благодаря усиленному действию задней перстнечерпаловидной мышцы, принимает форму пятиугольника.

Необходимо отметить, что механизм голосообразования представляется чрезвычайно сложным процессом, и попытки изучить его предпринимаются учеными давно [38,77,108,125,127]. Однако до настоящего времени этот вопрос нельзя считать разрешенным во всей полноте [37,57,58,59,60,90,117,142,143].

Поскольку длина и ширина истинных голосовых складок всегда привлекают внимание фониатров, на этих параметрах голосовых складок следует остановиться отдельно. По данным авторов [38], длина истинных голосовых складок как у мужчин, так у женщин колеблется от $\frac{1}{4}$ до $\frac{1}{3}$ высоты гортани и от $\frac{9}{11}$ до $\frac{11}{10}$ высоты угла щитовидного хряща. Что касается колебаний длины истинных голосовых складок по отношению к высоте дуги перстневидного хряща, то они у мужчин и женщин различны: у мужчин длина истинных голосовых складок колеблется в пределах от 2,5 до 3,1 высоты дуги перстневидного хряща, а у женщин — в пределах от $\frac{21}{5}$ до $\frac{24}{5}$ этой же величины. В 12—13 лет длина голосовых складок равна 13—14 мм; в период мутации длина их увеличивается на 6—8 мм, а к 25 годам достигает 22—25 мм. У девочек в переходном возрасте голосовые складки растут значительно медленнее, и длина их у взрослых достигает 18—20 мм [38].

Некоторые авторы при определении типа певческого голоса рекомендуют принимать во внимание антропометрические данные, так как наблюдения показывают, что обладатели низких голосов, как правило, отличаются более высоким ростом и хорошо развитой грудной клеткой. Наоборот, обладатели высоких голосов отличаются меньшим ростом и меньшим объемом грудной клетки. Так, например, в [38] приводятся следующие данные, которые представлены в таблице 1.3.

Таблица 1.3. Физиологические константы

Тип голосов	Средний рост, см	Длина речеобра- зующего тракта, см	Ширина голосовых складок, мм	Средний объем грудной клетки, см ³	Длина голосовых складок, мм
Тенор	166,2	15–17	2,5–3	96,0	18–22
Баритон	173,7	17–20	3–3,5	99,7	20–24
Бас	177,9	19–20	3,4–3,5	98,4	24–25

По данным, систематизированным в работе [117], резонансные частоты колебаний ненапряженных голосовых складок для мужских голосов лежат в пределах 91 – 145 Гц и для женских 115 – 145 Гц. При этом добротность голосовых складок составляет для мужских голосов 2,1 – 5,1, для женских 2,6 – 7,1, что соответствует коэффициенту затухания 0,08 – 0,25.

По-видимому, эквивалент голосовых складок для пищеводного голоса характеризуется близкими параметрами [2, 88, 121].

1.5 Модели голосообразования

История исследований процессов речеобразования при синтезе речи базируются на двух подходах: акустической теории Гуннара Фанта [125] и компилятивной теории. В результате в 1982 году Лобанов Б.М. создал синтезатор человеческой речи 15КС2000-014 [75], послуживший прототипом Фонемофона[77]. Деннис Клатт [149] в 1986 году использовал эти методы для языков романо-германской группы. Рассмотрим подход физиологов [38]. Однако, как уже отмечалось, синтез по правилам не дал положительных результатов.

Это связано с тем, что модели генерации звучной речи, по-видимому, не были адекватными процессам голосообразования в реальной системе.

Резонансную природу образования речевых гласных звуков обосновал еще Гельмгольц Г. в 1860 году созданием «говорящей машины» [84]. К этому времени относится возникновение так называемой *миоэластической* теории фонации [37,38,39].

Основными факторами голосообразования, согласно этой теории, является давление воздуха в трахее. Высота основного тона зависит от частоты колебаний голосовых складок, укорочения их звучащих отрезков и степени напряжения голосовых мышц, а интенсивность звука — от силы выдыхаемой струи воздуха. Сложная система дыхательных мышц рефлекторно поддерживает трахеальное и бронхиальное давление на определенном уровне, необходимом в связи с различными условиями голосообразования. По мере повышения и усиления звука подскладочное давление увеличивается, и наоборот. Благодаря этому же механизму при речи и пении обеспечивается возможность длительного (от 25 до 40 с и более) использования голоса без повторных вдохов. По данным некоторых авторов [38], в процессе голосообразования принимает участие и гладкая мускулатура трахеи и бронхов.

В противовес *миоэластической* теории колебаний голосовых складок французский исследователь Рауль Юссон (1960, 1962, 1965) [134, 147,148] выдвинул *нейрохроноксическую* теорию. Согласно этой теории, голосовые складки человека колеблются не пассивно под действием потока воздуха, а активно — периодически сокращаются и расслабляются со звуковой частотой, регулируя тем самым прохождение через гортань воздушных импульсов и, следовательно, частоту основного тона звука. В пользу этой теории говорят полученные факты о том, что во время фонации гласных (у ларингоэктомированных больных) частота основного тона голоса и частота нервных импульсов, идущих к голосовым связкам по двигательному нижнегортанному нерву, полностью совпадают [134, 147].

По поводу своей концепции механизма фонации Юссон [134] пишет следующее: “То, что издавна неосновательно называли “колебаниями” голосовых складок, не является в строгом смысле колебаниями, возникающими в упругих телах: это просто серия сверхкоротких и быстрых ритмических сокращений гертлеровских волокон голосовых складок”.

Данная теория колебания голосовых складок вызвала среди специалистов большую дискуссию [125, 134, 147]. Одно из основных возражений, выдвигаемых противниками Юссона, состоит в том, что передача высокой частоты раздражения к голосовым складками 1000 — 2000 Гц не может быть обеспечена ввиду наличия рефракторной фазы нерва. Как установлено классическими исследованиями [38], лабильность нерва не превышает 400—500 Гц. Особые свойства синаптической передачи приводят к тому, что функциональная подвижность системы “нерв — мышца” оказывается еще более низкой [3,103].

Отвергая это возражение, Юссон, во-первых, ссылается на особые свойства голосовых складок человека [38], которые, по его мнению, обладают значительно более высокими функциональными возможностями. Во-вторых, на наличие у человека особого механизма функционального деления нервных окончаний на части, по каждой из которых передаются импульсы с частотой, не превышающей нормальную лабильность нерва, но ввиду смещения по фазе обуславливающую суммарную частоту сокращений голосовых складок, в 2, 3 и даже 4 раза более высокую, чем частота импульсов, идущих по каждой из отдельных частей. Таким образом, если лабильность отдельного волокна составляет 400 Гц, то в общей сложности нерв способен передать возбуждение с частотой до 1600 и даже до 2000 Гц, если предположить возможность его функционального деления на пять частей, по каждой из которых передаются импульсы с опозданием на $\frac{1}{3}$ периода.

Основываясь на своей нейромоторной концепции механизма колебаний голосовых складок, Юссон считает, что можно определить предельные возможности частоты колебаний голосовых складок человека путем измерения

хронаксии нерва, возбудимость которого, как полагает Юссон, соответствует возбудимости голосовых складок. Юссон [134] показал, что наиболее низкие типы голосов (басы, контральто) имеют скорость воздушного потока 120-170 м/с, а высокие типы голосов (тенора, сопрано) 55—70 м/с. Поскольку правильное определение тембра голоса имеет большое практическое значение для успешности обучения певца, метод, который предложил Юссон, нашел применение в некоторых консерваториях Франции.

Одним из главных выводов этой теории является абсолютная независимость частоты колебаний голосовых складок от воздушного подскладочного давления, создаваемого в легких. «Подскладочное давление не может оказывать влияние на высоту звука, которая зависит исключительно от частоты рекуррентных импульсов», - писал Юссон [147]. Он считает, что силы нейромоторного происхождения, заставляющие голосовые складки колебаться, примерно в 10 раз больше сил миоэластического происхождения, связанных с воздушным давлением. Это приводит к полной независимости частоты колебаний голосовых складок от силы подскладочного давления воздуха. Последнее, как полагает Юссон, может оказывать влияние лишь на силу голоса, но не на его высоту. Этот основной вывод Юссона является необоснованным [128].

Колеблющиеся голосовые складки человека в нормальных условиях фонации не безразличны к изменению величины подскладочного давления, а довольно определенно реагируют на него [93], увеличивая частоту своих колебаний при увеличении давления и уменьшая число колебаний при ослаблении подскладочного давления. Явление это можно объяснить с позиций *миоэластической* теории, так как при повышении подскладочного давления складки рефлекторно смыкаются плотнее и более напрягаются, чтобы противостоять усилившемуся давлению. Это обстоятельство приводит к увеличению упругих свойств голосовых складок (увеличивается жесткость колеблющейся системы), что, в свою очередь, и ведет к увеличению собственной частоты колебаний складок. Из теории колебания упругих тел

известно, что частота колебаний пропорциональна жесткости колеблющейся системы и величине воздушного давления, приводящего тело в колебательное состояние. Однако в нормальных условиях фонации голосовые складки не предоставлены самим себе, как в изолированной гортани, а постоянно находятся под контролем центральной нервной системы, регулирующей частоту и силу их колебаний. Поэтому, как только изменение подскладочного давления вынуждает голосовые складки повысить или понизить частоту колебаний, центральная нервная система “принимает срочные меры” к тому, чтобы восстановить прежнюю частоту их колебаний. Меры эти, в сущности, сводятся к компенсаторному изменению эластических свойств голосовых складок путем соответствующего перераспределения их натяжения, плотности, жесткости и т. д.

Легко проверить корректирующую роль слухового анализатора, “выключив” его из участия в контроле за высотой производимого голосовым аппаратом звука. Последнее достигается подачей на уши испытуемого (через головные телефоны) сильного шума (до 110 Дб) от шумового генератора. В этих условиях регулирующая обратная акустическая связь прерывается, испытуемый перестает слышать свой голос, что приводит к значительно большим, чем в норме, изменениям высоты основного тона при периодическом изменении силы голоса.

Установлено [38, 121], что повышение давления в ротовой полости во время фонации приводит к закономерному прорежению частоты колебаний голосовых складок, а также к уменьшению амплитуды их вибрации. У одного из испытуемых со слабым нетренированным голосовым аппаратом повышение давления в ротовой полости до 200 мм вод. ст. вызвало не только явные прорежения колебаний голосовых складок, но и аритмию и почти полное прекращение колебаний. После спада давления в ротовой полости до уровня нормального частота колебаний голосовых складок не только быстро вновь повышается до нормы 200 Гц, но и значительно превосходит эту норму доходя до 220—230 Гц, потом снова уменьшается ниже нормы и, наконец,

устанавливается на стабильном уровне. Восстановление частоты колебаний голосовых складок происходит как бы волнообразно, постепенно приближаясь к норме. Повышение давления в ротовой полости менее чем на 100 *мм вод. ст.* заметно меньше сказывается на частоте колебаний голосовых складок.

Таким образом, повышение давления в ротовой полости приводит к такому же эффекту (понижению частоты колебаний складок), что и уменьшение подскладочного давления. Это свидетельствует о том, что голосовые складки реагируют на градиент воздушного давления.

Кроме того [117], установлено, что подскладочное давление, т.е. давление на выходе трахеи, при раскрытии голосовой щели может быть в 5 – 6 раз меньше максимального. Следовательно, можно ожидать, что колебания голосовых складок согласованы с изменениями давления на выходе трахеи. Считается [117], что на каждые 10 *мм. вод. ст.* частота основного тона увеличивается на 2 – 4 *Гц*.

Исследования показали [38, 121], что реакция голосовых складок проявляется не столько на абсолютную величину давления, сколько на скорость его нарастания или ослабления: чем больше скорость нарастания (или уменьшения) давления, тем заметнее замедляются (или учащаются) колебания складок, и наоборот — складки мало реагируют на медленно возрастающее или плавно спадающее давление. Явление это, по-видимому, обусловлено известной инертностью рефлекторных механизмов, поддерживающих высоту основного тона на определенном уровне: если время, в течение которого голосовые складки вынуждаются под действием давления изменить частоту своих колебаний, достаточно велико (скорость изменения давления $\Delta P/t$ мала, где t — время), то центральная нервная система оказывается в состоянии обеспечить контроль за частотой колебаний складок и поддерживать эту частоту на определенном уровне. Если же скорость изменения градиента давления велика, то частота колебания голосовых складок изменяется раньше, чем рефлекторный механизм коррекции успевает вступить в действие.

Приведенные экспериментальные данные [38] свидетельствуют о том, что градиент воздушного давления на уровне голосовых складок является и в условиях живого организма весьма существенной силой, влияющей на частоту колебаний голосовых складок. Таким образом, то, что Юссон [134] абсолютно отрицает действенность этой силы, является неправомерным.

Из вышеизложенного вытекает, что механизм регулирования частоты колебаний голосовых складок имеет сложную биофизическую природу. Немаловажную роль в этом процессе играет градиент воздушного давления на уровне голосовых складок. По данным Фанта (1964) [125], частота основного тона голоса при прочих равных условиях пропорциональна квадратному корню подскладочного давления.

Однако сила подскладочного давления — не единственная причина, заставляющая голосовые складки изменять частоту вибрирования. Многочисленные опыты на изолированной гортани, а также стробоскопические наблюдения на живом человеке позволили убедиться, что частота звука, издаваемого гортанью, обратно пропорциональна длине колеблющейся части голосовых складок [38,121]. Эта закономерность находит отражение, в частности, в том, что у высоких типов голосов средняя длина голосовых складок в 1,5 — 2 раза короче, чем у низких см. табл. 1.3. Длина колеблющейся части голосовых складок у каждого певца является произвольно регулируемым параметром благодаря сложному синцитиальному переплетению мышечных волокон. Кроме длины колеблющейся части, частота вибраций голосовых складок (F_0) определяется также их толщиной, точнее — массивностью и жесткостью в соответствии с формулой, т.е. обратно пропорциональна квадратному корню из массы колеблющейся части [38].

Жесткость складок регулируется в широких пределах путем мускульного напряжения. Массивность колеблющейся части также может регулироваться в известных пределах. Так, например, стробоскопическими и томографическими исследованиями установлено, что в низком (“грудном”) регистре голосовые складки колеблются всей своей массой, а в высоких регистрах (“головном” и

“фальцетном”) вибрирующая область ограничивается краями голосовых складок.

Из теории колебаний известно, что на частоту вибратора сильное влияние может оказывать прилегающий к нему акустический резонатор [38,125]. Влияние этой закономерности, давно нашедшей свое отражение в конструкции многих музыкальных инструментов (деревянные духовые и язычковые органнне трубы), обнаружено в начале 70-х годов и в принципах работы голосового аппарата человека. Дмитриевым Л. Б. [37] согласно [38] было показано, что длина надставной трубки (верхних резонаторов) певцов тем больше, чем ниже голоса. Измерения длины резонаторов (по рентгеновским снимкам) приведены в табл. 1.2. (по данным [38]).

Исследуя речь человека в искусственных атмосферах, т.е. в космосе и при погружении используют специальное снаряжения на большие глубина, где основной составляющей дыхательной смеси используется азот, наблюдаются следующие результаты [79, 93]: во-первых, частотный спектр смещается в область высоких частот, при сохранении индивидуальных особенностей диктора, во-вторых, укорачивается время вдоха и выдоха.

Таким образом, доказано, что на частоту колебаний голосовых складок может влиять целый ряд факторов: их длина, массивность, жесткость, сила подскладочного давления, объем и форма резонаторов и прочее. Юссон считает [147], что роль всех этих сил в регулирования частоты колебаний голосовых складок пренебрежительно мала по сравнению с обнаруженными им силами прямого *нейромоторного* происхождения. Как нам кажется, приведенные данные не подтверждают точку зрения Юссон [147], а скорее свидетельствуют о том, что все многообразие форм колебаний голосовых складок можно объяснить, исходя из учета всех вышеуказанных факторов биофизического происхождения. В числе последних следует указать также на весьма важную роль *эффекта Бернулли*, возникающего при движении частиц воздуха между краями голосовых складок и поддерживающего их автоколебательное состояние (Г. Фант, 1964) [125].

Этими доводами, однако, отнюдь не умаляется роль центральной нервной системы в регулировании частоты колебаний складок. Напротив, следует подчеркнуть, что все эти акустико-механические факторы (длина колеблющейся части, жесткость, подскладочное давление, длина резонаторов и т. д.) регулируются центральной нервной системой в соответствии с теми или иными задачами звукообразования. Юссон [134] указывает, что роль нервной системы состоит в непосредственном ритмическом сокращении мускульных волокон вибратора. По нашему мнению, нервная система регулирует частоту вибратора более сложным опосредованным путем, т. е. путем создания таких акустических и механических условий для вибратора, при которых он порождает звук определенной частоты. При этом огромную роль имеет система биологических обратных связей на основе функции слухового, вибрационного и проприорецептивного анализаторов, сигнализирующих в центральную нервную систему, как о состоянии органов звукопроизношения, так и о результате их деятельности (Морозов В. П., 1965) [38,85].

Вместе с тем, наличие не одного, а целой системы перечисленных выше биофизических механизмов регулирования голосовых складок человека создает возможность их взаимной компенсации и замещения, что обеспечивает не только высокую точность и надежность регулирования основного тона, но (при неизменном основном тоне) позволяет варьировать в известных пределах и другие важнейшие акустические параметры звука: его тембр и силу голоса. Этим, в частности, объясняется, почему тренированные певцы способны изменять в определенных границах силу и тембр голоса, сохраняя при этом высоту основного тона относительно постоянной. Этим же объясняется возможность фонации при заболевании голосового аппарата и расстройствах некоторых регуляторных механизмов: другие механизмы берут на себя их функцию, т. е. компенсируют утраченные механизмы.

От частоты колебаний голосовых складок зависит основной тон голоса [38]. Последний, является важным информационно несущим параметром речевого сигнала, так как определяет слуховое восприятие высоты звука.

Следует, однако, заметить, что на восприятие высоты звука известное влияние оказывает также и спектральный состав сложных звуков. Звуки, богатые высокими спектральными составляющими, воспринимаются на слух как более высокие по сравнению со звуками, имеющими ту же частоту основного тона, но со слабо выраженными высокими гармониками.

Высокочастотные составляющие речевого (для русской речи принято считать данный диапазон начиная с 1200 Гц, т.е. со второго спектрального максимума [108, 127]) сигнала несут слушателю информацию различного рода. Во-первых, ими обуславливается интонация предложения (вопрос, восклицание, повествование), во-вторых — ударение, в-третьих — смысловое значение различных частей речи, что особенно выражено в так называемых тональных языках (японский, индонезийский и др.), и, в-четвертых — эмоциональное отношение говорящего к предмету разговора.

Изменение частоты основного тона является настолько характерным признаком нормальной речи, что отсутствие высокочастотных составляющих спектра речи воспринимается как резкое нарушение ее естественности. Согласно [38] считается, что отсутствие изменений основного тона речи или нарушение его характерных особенностей может служить признаком психической ненормальности говорящего. Подобное неприятное впечатление нередко возникает при слушании синтетической речи, лишенной естественной интонации (“речь роботов”). Этот недостаток в значительной мере устраняется путем имитации характерных изменений частоты колебаний основного тона, искусственного речевого сигнала. Спектральный состав человеческой речи приведен в табл. 1.4.

Таблица 1.4 Звуковой диапазон основных типов голосов[38]

Мужские голоса	В нотном обозначении	Основной тон в, Гц	Длина средняя, см	Женские голоса	В нотном обозначении	Основной тон в, Гц	Длина средняя в, см
Нормаль	136 Гц	85-200	17,5	Нормаль	248 Гц	160-340	16,7
Бас	mi-fa'	82-349	22,3-25	Контральдо	mi-fa''	165-698	-
Баритон	la-sol'	110-932	-	М-сопрано	La-sol''	220-880	-
Тенор	bo-do''	131-523	19-22	Сопрано	Bo-do'''	262-1046	15,3-18,5

Таким образом, формирование речи существенно выходит за рамки одного нейрорхонкстического или милоэластического подхода. К тому же необходимость описания пищеводной речи заставляет нас найти общие части в акустических моделях, представленных в табл. 1.1. В дальнейшем сформированная нами модель для нормальной речи могла бы быть с минимальными изменениями трансформирована в модель пищеводной речи.

Можно отметить:

- милоэластическая гипотеза основана на том, что голосовые складки раскрываются под действием давления в трахее и смыкаются под действием сил Бернулли;
- нейрорхонкстическая гипотеза основана на том, что на движение голосовых складок не влияет давление как в трахеи, так и в ротовой полости.

Практически эти гипотезы дали основание Фанту Г. [125] высказать предположение, что голосовые складки могут рассматриваться как источник тока или напряжения. Следовательно, движение голосовых складок и процессы в речеобразующем тракте не влияют друг на друга.

Однако данные о зависимости частоты основного тона от давления в ротовой полости [38, 121] и в трахее [93, 117] позволяют предположить, что процессы в трахеи, в ротовой полости и движение голосовых складок взаимосвязаны.

1.6 Постановка задачи исследования

Традиционно генерация вокализованных звуков речи представлялась схемой показанной на рис. 1.4, которая не противоречит нейроронаксической и мелозластической гипотезам.

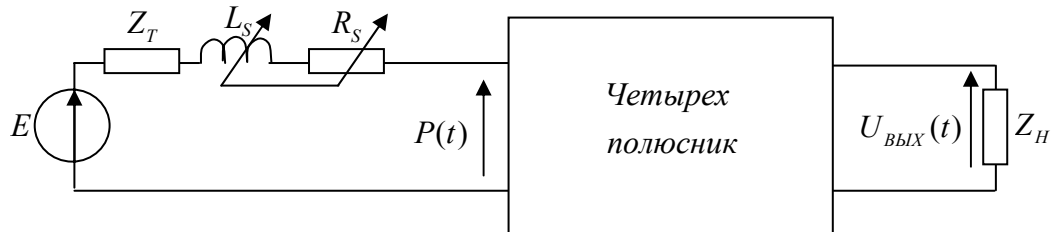


Рисунок 1.4 Традиционная модель голосообразования

В основу этой схемы положена гипотеза, высказанная Фантом [125], о том, что функционирование различных частей речеобразующей системы можно рассматривать независимо друг от друга, т.е. справедливы соотношения:

$$|Z_S| \gg Z_{BX}, \quad |Z_T| \ll |Z_S|, \quad (1.1)$$

где Z_S – выходной импеданс голосовой щели; $Z_S = sL_s + R_s$;

L_s, R_s – реактивное и активное сопротивления голосовой щели;

Z_{BX} – входной импеданс четырехполюсника;

Z_T – выходной импеданс трахеи;

E – эквивалентное емкости легких напряжение.

Таким образом, если $P(j \cdot \omega)$ – давление на выходе складок, а $H(j \cdot \omega)$ – передаточная функция четырехполюсника, тогда

$$U_{ВЫХ}(j \cdot \omega) = P(j \cdot \omega) \cdot H(j \cdot \omega). \quad (1.2)$$

Ясно, что в этой модели форманты формируются функцией $H(j \cdot \omega)$. Данный подход моделирования показал, что даже при использовании очень сложных моделей голосовых складок [108,112,115] получается разборчивая, но имеющая явно выраженный механический оттенок речь. Таким образом, можно предположить, что гипотеза Фанта о независимом функционировании

отдельных частей речеобразующей системы неверна. Учет возможных взаимосвязей приводит к схеме, представленной на рисунке 1.5.

Исследование механизмов образования речи на основе этой схемы предполагает совместное рассмотрение процессов, происходящих в легких C_L , трахее Z_T , голосовых складках L_S , R_S , надгортанной области Z_{POT} , Z_H . В этом случае формантная структура речи формируется сразу во всей системе.

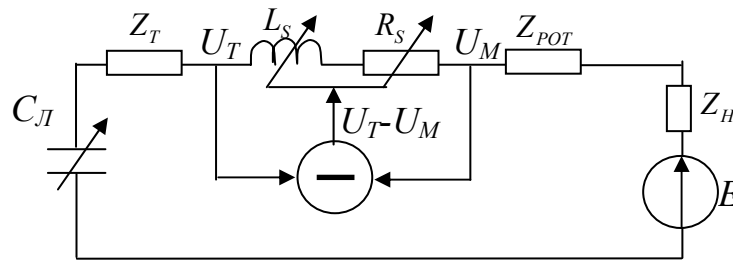


Рисунок 1.5 Схема электрического аналога речеобразующей системы.

Порция воздуха втягивается в легкие и накапливается в их акустической емкости C_L . Уменьшение C_L повышает легочное давление, которое распространяется по бронхам и трахее до голосовых складок. Таким образом, создается разность давлений в трахее и морганьевом (гортанном) желудочке. Разность давлений $U_T - U_M$, в свою очередь, вызывает ток воздуха i_C через относительно узкое отверстие между голосовыми складками. Под воздействием изменений местного давления в голосовой щели складки, обладающие определенной массой и эластичностью, приходят в колебательное движение. Квазипериодические размыкания и смыкания складок приводят к колебаниям импеданса $Z_S = (R_S + i\omega L_S)$ и модулируют воздушный поток. Поэтому поток воздуха, проходящий в голосовой тракт, имеет форму дискретных импульсов.

В схеме на рис. 1.4 считается, что изменение импеданса Z_S голосовой щели обратно пропорционально площади раскрытия голосовых складок, т.е. проводимость щели изменяется от нуля до максимального значения, и что колебания складок не зависят от свойств последующих органов. Однако следует ожидать, что существенное влияние на колебания голосовых складок оказывает входной импеданс голосообразующего тракта. Необходимо отметить, что есть экспериментальные данные [30], которые говорят о том, что

речевой сигнал полностью оформлен на выходе голосовой щели. Это подтверждает существование связи колебаний складок с входным импедансом речеобразующего тракта. Возможность использования стандартной модели речеобразования, предложенной Фантом, существует, но необходимо учитывать самовозбуждение системы речеобразования, на участке голосовых складок R_S и L_S , структура которой приведена на рис. 1.5.

Упрощенную модель образования нормальной речи можно представить в виде схемы, изображенной на рис. 1.6.

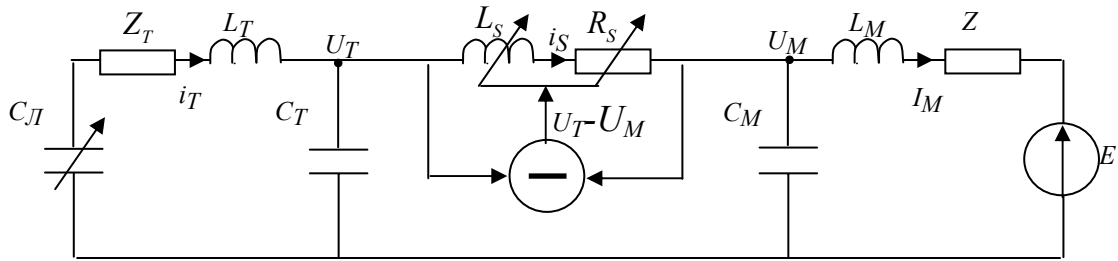


Рисунок 1.6 Уточненная модель речеобразования

На рис. 1.6 использована простейшая модель легких и трахеи, рассмотренная в работе [116] и построенная с учетом [81, 89, 128], где R_T , L_T и C_T - потери, индуктивность и емкость электрического аналога трахеи и легких в модели [117]; L_M , C_M и R_M - индуктивность, емкость и потери электрического аналога морганьевого желудка; Z – входной импеданс речеобразующего тракта.

Из этой модели следует, что:

$$i_S(t)R_S(t) + \frac{d(L_S(t)i_S(t))}{dt} + i_S^2(t)\lambda = U_T - U_M, \quad (1.3)$$

$$U_T = \frac{1}{C_T} \int i_T(t) dt - \frac{1}{C_T} \int i_C(t) dt, \quad (1.4)$$

$$U_M = \frac{1}{C_M} \int i_C(t) dt - \frac{1}{C_M} \int i_M(t) dt. \quad (1.5)$$

В формуле (1.3) квадратичная компонента тока в левой части уравнения учитывает влияние сил Бернулли, λ – коэффициент пропорциональности. Зависимость индуктивности и емкости голосовых складок от ее площади раскрытия описывается выражениями [125, 126]:

$$L_s = \frac{\alpha}{S}, \quad R_s = \frac{\beta}{S}, \quad (1.6)$$

где S – площадь раскрытия голосовых складок;

α и β – коэффициенты пропорциональности.

Таким образом, колебания связок, т.е. характер зависимости R_s, L_s от времени, будет определяться разностью $U_T - U_M$, что может привести к автосинхронизации колебаний голосовых складок с процессами, происходящими в трахее и голосообразующем тракте. Необходимо учитывать, что характер движения складок, т.е. вид функций $R_s(t)$ и $L_s(t)$, описывается нелинейным уравнением второго порядка в частных производных [117], с правой частью, определяемой величиной $U_T - U_M$, которая зависит от входного импеданса Z . В простейшем случае движение голосовых складок в модели рис. 1.6 можно описать дифференциальными уравнениями второго порядка:

$$\frac{d^2 S}{dt^2} + \delta \frac{dS}{dt} + \omega^2 \cdot S = \frac{U_T - U_M - \gamma \cdot i_s^2}{a}, \quad (1.7)$$

где δ – потери в голосовых складках;

ω – собственная частота колебаний голосовых складок;

γ – коэффициент пропорциональности, учитывающий силы Бернулли;

a – коэффициент пропорциональности.

Таким образом, для исследования модели, представленной на рис. 1.6 необходимо получить аналитическое выражение для Z . За основу Z_T , в первом приближении, можно использовать модели, рассмотренные в работе [116], в частности модель, представленную электрическими аналогами цепи второго порядка. Кроме того, представленная модель требует дополнительного анализа с целью определения условий возникновения устойчивых колебаний, так как в ней в явном виде присутствует обратная связь.

1.7 Адекватность модели

Оценка адекватности модели в данных исследованиях сопряжена со многими трудностями. Ясно, что полного соответствия модели речеобразующей системы человека в настоящее время достичь практически

невозможно. Это связано с неполными и зачастую недостоверными данными по анатомии, физиологии и нейродинамики отдельных частей и всей системы речеобразования в целом. Поэтому за основу оценки адекватности модели возьмем метод «черного ящика». То есть будем оценивать соответствие модели по результатам сравнения сгенерированных сигналов с реальными.

Такое сравнение возможно проводить двумя методами:

- сравнение их динамических спектрограмм [32, 34, 36, 83, 110, 113];
- субъективной оценкой качества (разборчивости, естественности и узнаваемости) сгенерированных речевых сигналов [25, 26, 27, 96].

Известно [28, 31], что в линейных и параметрических системах, а рассматриваемая модель относится к таким системам, возникает достаточно много спектральных компонентов, находящихся в определенных амплитудно-фазовых соотношениях между собой. Поэтому обычные методы спектрального анализа при сопоставлении сгенерированных и реальных речевых сигналов могут оказаться неприемлемыми. Для такого анализа использовались методы синхронного с периодом основного тона спектрального анализа, разработанного при непосредственном участии Пономарева А.А. [19, 20, 54].

Оценка естественности, узнаваемости и разборчивости речевых сигналов, как правило, ведется аудиторными испытаниями.

Трансформация амплитудного спектра, положение максимумов в спектре речевого сигнала мало сказываются на восприятие соответствующего звука речи. До сих пор нет приемлемого синтезатора речи, который бы удовлетворил не только неподготовленных пользователей, но и специалистов. Примеры программ, синтезирующих человеческую речь, и их анализ приведены в [33, 57, 61, 77].

Приходится констатировать, что предложенные до сегодняшнего дня методы решения волнового уравнения всеми описанными выше методами не проходили испытания на ЕСТЕСТВЕННОСТЬ и РАЗБОРЧИВОСТЬ. Эта процедура необходима для оценки качества синтезированной речи.

Теория разборчивости речи [25, 26, 27] основывается на статистической теории приёма сигналов. Именно поэтому между теорией информации и теорией разборчивости существует тесная связь.

Не останавливаясь на изложении основ теории разборчивости, которые подробно освещены в работах [26, 27, 29, 116], рассмотрим их только в аспекте основных положений теории информации. Само собой разумеется, что при таком рассмотрении не надо забывать того, что приёмником речевого сигнала является (в этом случае) слух человека, имеющий характеристики, сильно отличающиеся от характеристик, обычно используемых приёмников электрических сигналов [124, 133].

Следующим моментом, который необходимо учитывать при анализе условий передачи речи, является разница в оценке её: в одних случаях речь оценивают только с точки зрения её понятности, а в других — с точки зрения и разборчивости и качества звучания. Промежуточным случаем может явиться такой, при котором, кроме понятности речи, необходима и узнаваемость голоса. Конечно, между этими оценками имеется связь, и в случае непреобразованной речи эта связь почти без исключений однозначная. Для преобразованной речи нельзя утверждать, что такая связь однозначна, но всё же довольно часто пытаются оценивать качество передачи речи только по её разборчивости. В дальнейшем на этом вопросе придётся остановиться несколько подробнее в связи с рассмотрением конкретных устройств преобразования речи. В данном разделе под понятием качества передачи речи мы будем подразумевать качество разборчивости речи.

Исследования, проведённые ВКАС и рядом НИИ [26, 25, 27], по определению зависимости между качеством передачи речи и её разборчивостью на различных линиях связи показали, что для непреобразованной речи эта зависимость однозначна. Для введённых оценок “отлично”, “хорошо”, “удовлетворительно” и “предельно допустимо” в результате массовых испытаний (участвовало несколько сотен слушателей с разнообразной подготовкой и развитием) было установлено, что абсолютное

большинство слушателей (не менее 90%) давало оценку “отлично” при словесной разборчивости не менее 98%, оценку “хорошо” - не менее 93%, оценку “удовлетворительно” - 87%, оценку “предельно допустимо” - 75%. Кроме того, профессором Быковым [26] была введена оценка “срыв связи”, которая соответствовала словесной разборчивости (по данным ВКАС) 60%. Если к этим градациям качества передачи речи добавить ещё оценку “идеальная передача”, под которой подразумевать такую, при которой ни один из слушателей не замечает искажений, то по словесной формантной [26] разборчивости будем иметь пределы указанные в таблице 1.5.

Таблица 1.5 Критерии оценки качества синтезированной речи

	Словесная разборчивость (W)	Слоговая разборчивость (S)	Формантная разборчивость (A)	Естествен- ность
идеально	100—99%	100—75%	100—60%	100%
отлично	>97(95)%	>75%	>60%	
хорошо	>94(90)%	>60%	>52%	38%
удовлетворительно	>98(80)%	>45%	>46%	
предельно допустимо	73(80)%	>32%	>42%	
срыв связи	<73—65%	<20%	<36%	

Термин естественность или натуральность использовался для оценки вокодеров, со временем его стали применять для определения качества синтетической речи. Методология была использована из [27], которая предполагала следующие ограничения: полоса пропускания от 120 до 3500 Гц, количество ступеней при разбиении диапазона 13,5.

«В западной литературе термин естественность для синтезаторов речи по правилам определяется через соотношение сигнал-шум», - пишет Сорокин В.Н. в [116].

Основными методами синтеза речи являются метод компиляции и форматно-волновой, они не обеспечивают необходимые потребительские характеристики (разборчивость на уровне «хорошо», а естественность «неудовлетворительно»). Все оценки качества созданы на основе аудиторных испытаний. Чаще всего на базе слоговых таблиц. Поэтому необходимо

тщательно рассматривать процессы, происходящие в речеобразующем тракте, и, используя точные модели, синтезировать естественную речь человека.

Предварительная оценка адекватности модели проводилась на основе сравнения спектральных характеристик сгенерированных речевых сигналов, а окончательная оценка адекватности на основе субъективных критериев в соответствии с рекомендациями [25, 26, 27].

Выводы по главе:

Рассмотрена анатомия органов речеобразования при формировании нормальной и пищеводной речи. Составлены акустические схемы процессов голосообразования нормальной, пищеводной и фарингеальной речи, на основе которых выделены общие части этих речеобразующих систем.

Исследованы основные анатомические, физиологические, нейродинамические, акустические и механические параметры легких, бронхов, трахеи, воздушного пузыря, гортани, голосовых складок и псевдоголосовой щели. Установлены различия в рассматриваемых речеобразующих системах.

Проанализированы существующие модели голосообразования. Показано, что миэластическая и нейроронаксическая гипотезы не объясняют всех известных данных по механизмам голосообразования. На основе этого можно предположить, что эти гипотезы являются частными проявлениями общего механизма, ответственного за голосообразование.

Предложен возможный механизм голосообразования, который явился основной постановки задачи. Этот механизм предполагает, что колебания голосовых складок определяются их механическими свойствами и разностью давлений в трахее и ротовой полости. Тогда роль нервной активности сводится, в основном, к управлению собственной частотой колебаний голосовых складок. Но в этом случае возникает дополнительная задача, связанная с анализом устойчивости процессов в речеобразующей системе.

Изучены вопросы, связанные с установлением адекватности моделей речеобразования реальных систем. Основным критерием оценки предлагается использовать результаты сопоставительного анализа выходных сигналов, генерируемых в модели с реальными речевыми сигналами, также субъективные оценки качества синтезированных в модели сигналов.

ГЛАВА 2. МОДЕЛЬ ОБРАЗОВАНИЯ НОРМАЛЬНОЙ И ПИЩЕВОДНОЙ РЕЧИ

Модели голосообразования, как следует из постановки задачи, должны учитывать возможные нелинейные механизмы, вызванные процессами возбуждения, т.е. процессами, происходящими в системе на уровне голосовых складок или их эквивалентов.

Как уже отмечалось, движение складок и воздушный поток через них должны зависеть от процессов, происходящих в легких и трахее, а также в голосообразующем тракте. Если для легких и трахеи уже имеются некоторые модели [125,127], представленные в виде электрических аналогов, то для речеобразующего тракта имеющиеся модели требуют решения весьма сложных уравнений в частных производных. Поэтому первым этапом построения адекватной модели является разработка точного электрического аналога речеобразующего тракта, который позволял бы производить численное моделирование процессов голосообразования. Естественно, что данная модель речеобразующего тракта должна базироваться на уже известных моделях, а также учитывать известные свойства и параметры реального речевого сигнала.

2.1 Входной импеданс речеобразующего тракта

Представление речевого тракта в виде эквивалентной электрической схемы основывается на понятии акустического импеданса [125], определяемого как отношение звукового давления к объемной скорости. Его не следует смешивать с удельным импедансом, представляющим собой отношение давления к скорости частиц, а также с механическим импедансом, равным отношению силы к скорости частиц. Преимущество использования именно акустического импеданса определяется непрерывностью изменения как объемной скорости, так и звукового давления в пределах всей системы резонирующих полостей.

Первая попытка формализации процессов речеобразования была предпринята с использованием основного закона колебания воздуха в узких трубах, которое описывается уравнением Вебстера:

$$-\frac{1}{S} \frac{\partial}{\partial x} \left(S \frac{\partial p}{\partial x} \right) = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2},$$

где $S=S(x,t)$ - площадь сечения трубы;

$p=p(x,t)$ - давление.

Данное уравнение, описывающее колебания воздуха в трубах переменного сечения, для нашей задачи не имеет пока аналитического решения. Одним из первых, кто воспользовался для описания процессов в речеобразующем тракте данным уравнения, был Фант [125]. Он качественно описал процессы в речеобразующем тракте при условии, что сечения не зависят от времени, но в дальнейшем воспользовался соответствующим электрическим аналогом, широко используемым для описания акустических процессов [72, 117]. Предположение о независимости сечений от времени допустимо для речеобразующего тракта на стационарном участке произнесения продолжительных звуков. Однако оно не применимо для участка, отвечающего за формирования импульсов основного тона. Поэтому в дальнейшем только для речеобразующего тракта воспользуемся результатами моделирования Фанта [125].

Математическое рассмотрение процесса речеобразования распадается на ряд последовательных операций. Первой из них является представление конфигурации речевого тракта в виде функции, характеризующей изменение площади поперечного сечения речевого тракта вдоль его оси, начиная от голосовой щели и до излучающего отверстия рта. Второй операцией является приближенное представление этой функции системой по возможности малого количества отрезков, в пределах которых площади можно считать постоянными. Далее, для этой системы находится функция передачи, которая умножается на частотную характеристику источника. Все это справедливо для случая, когда источник независим от речеобразующего тракта. При рассмотрении гласных звуков наибольшее значение имеют частоты первых пяти формант [108, 126,127].

При построении электрических аналогов, рассмотренных акустических схем необходимо учитывать соотношения между размерами речеобразующих систем l и длиной волны λ колебательных процессов в них [48]. «Если $l < \lambda$, то система имеет характер системы с сосредоточенными параметрами, если $l > \lambda$, реально много больше, то система приобретает свойства системы с распределенными параметрами. В частности, при максимальной длине голосообразующего тракта 0,2 м и скорости звука 350 м/с имеем, что для частот до 1750 Гц $l < \lambda$ и только для частот больше 1750 Гц $l > \lambda$. Практически всегда процессы генерации импульсов основного тона, т.е. процесс фонации, а также формирование формант, можно рассматривать на основе электрических аналогов с сосредоточенными параметрами» - писал Сорокин В.Н. [117].

Если рассматривать речеобразующий тракт как каскадное соединение четырехполюсников, то можно представлять его в виде большого числа последовательно соединённых электрических звеньев, каждое из которых является эквивалентом небольшого участка речевого тракта. Такой участок вполне можно рассматривать как короткий цилиндр с поперечным сечением, равным поперечному сечению речевого тракта на заданном расстоянии от гортани. В монографии Фанта [125] представлен один из таких цилиндров и его электрические аналоги для осевых колебаний. Каждый из электрических аналогов, имитирующих речевую секцию, можно представить в виде Т – (звеньев см. рис 2.1).

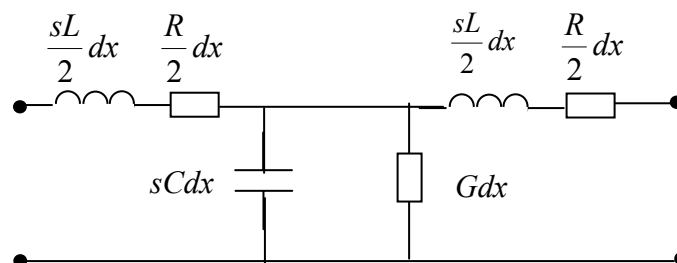


Рисунок 2.1 Электрический аналог секции трубы длиной dx

Каждое звено характеризуется, как четырех полюсник [41], A - параметрами, которые для модели с распределенными параметрами имеют вид:

$$\begin{aligned}
A_{11}^i &= A_{22}^i = ch(\gamma_i dx), \\
A_{12}^i &= Z_i sh(\gamma_i dx), \\
A_{21}^i &= \frac{sh(\gamma_i dx)}{Z_i},
\end{aligned} \tag{2.1}$$

а для модели с сосредоточенными параметрами:

$$\begin{aligned}
A_{11}^i &= A_{22}^i = 1 + \frac{1}{2}(sL_i + R_i)(sC_i + G_i) \cdot d^2 x, \\
A_{12}^i &= (sL_i + R_i) \left[1 + \frac{1}{4}(sL_i + R_i)(sC_i + G_i) d^2 x \right] \cdot dx, \\
A_{21}^i &= (sC_i + G_i) \cdot dx,
\end{aligned} \tag{2.2}$$

где L_i , C_i , R_i , G_i – погонная индуктивность, емкость, сопротивление и проводимость электрического аналога i -го участка трубы длиной dx ;

$$Z_i = \sqrt{\frac{sL_i + R_i}{sC_i + G_i}}, \quad \gamma_i = \sqrt{(sL_i + R_i)(sC_i + G_i)}. \tag{2.3}$$

Тогда матрица A - параметров всего голосообразующего тракта, может быть определена как произведение

$$A = \prod_{i=1}^n A^i,$$

где n – число секций аппроксимации сечений голосообразующего тракта.

Данные A - параметры определяют свойства только голосообразующего тракта, в частности входной импеданс Z , который равен:

$$Z = \frac{Z_0 A_{11} + A_{12}}{Z_0 A_{21} + A_{22}}, \tag{2.4}$$

где Z_0 - входной импеданс излучателя.

Значения параметров R_i , L_i , C_i , G_i , приведены в табл. 2.1, заимствованные из работы [146], в которой проведено обобщение данных, полученных в работе [127]. Более подробно с физиологическими константами пригодными для моделирования, можно ознакомиться в [135, 136, 137, 138, 139, 160, 161, 164, 165, 166, 167, 168, 169].

Где r – радиус трубы, в см; l – длина сегмента, в см; ω – угловая частота; ρ_0 — плотность стенок, в г/см³; η – вязкость среды, в динах с/см²; a – площадь

поперечного сечения, в см^2 ; c – скорость звука, в см/с ; ν – коэффициент удельной теплоемкости; κ – коэффициент теплопроводность, $\text{кал}/(\text{см}\cdot\text{с}\cdot^\circ\text{C})$; и c_p = удельная теплоемкость при постоянном давлении, в $\text{кал}/(\text{г}\cdot^\circ\text{C})$.

Таблица 2.1

Параметр	Значение
Сопротивление	$R_i = \frac{2l}{\pi^3} \sqrt{\frac{\omega \rho_0 \eta}{2}}$
Индуктивность	$L_i = \frac{\rho_0 l}{A}$
Податливость (емкость)	$C_i = \frac{Al}{\rho_0 c^2}$
Проводимость	$G_i = 2\pi l \frac{\nu - 1}{\rho_0 c^2} \sqrt{\frac{k\omega}{2c_p \rho_a}}$

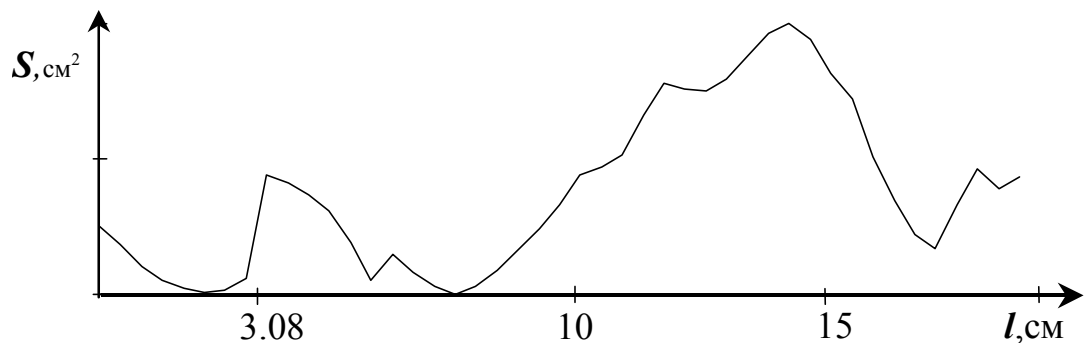


Рисунок 2.2 Зависимость площади поперечного сечения речеобразующего тракта от длины (голосовые складки-губы).

На рис. 2.2 приведена по данным работы [116] зависимость сечений речеобразующего тракта для звука «А» начиная от голосовых складок до губ. Для ряда других звуков зависимости приведены в приложении 1.

Необходимо выделить область глоточного желудочка сразу за голосовыми складками, после которых происходит резкий скачок сечения речеобразующего тракта. Поэтому можно считать, что выход голосовых складок нагружен на емкость и индуктивность глоточного желудочка, который, в свою очередь, нагружен на входной импеданс речеобразующего тракта. Поэтому будем

считать, что речеобразующий тракт начинается от указанного скачка сечений. Тем более что после резекции гортани глоточный желудок удаляется и нагрузкой эквивалента голосовых складок при пищеводной речи становится отрезок пищевода.

В табл. 2.2 приведены расстояния от голосовых складок до скачка сечений x_0 для разных звуков, а также общая длина речеобразующего тракта от складок до губ по данным, приведенным в приложении 1.

Таблица 2.2 Расстояние до сужения для разных звуков

Звук	А	О	У	И	Ы	Э	С	Ш	Ф	Х
x_0 (см)	3,08	3,08	3,52	1,76	3,96	3,08	1,76	3,08	3,08	3,08
l (см)	19,36	19,36	21,12	16,28	20,68	19,80	17,60	20,24	19,36	19,36

На основе соотношения (2.4), по данным табл. 2.1, табл. 2.2, и сечений приведенных в приложении 1, был проведен расчет входного импеданса речеобразующего тракта для разных звуков. На рис. 2.3, 2.4, 2.5 приведены результаты расчета импеданса для звука «А», а именно: амплитудно- и фазо-частотные характеристики, а также его действительная и мнимая часть.

Необходимо отметить, что расчет импеданса Z при использовании соотношений (2.1) и (2.3) дает практически одинаковые результаты. Это объясняется тем, что разница между соотношениями (2.1) и (2.2) при разложении (2.1) в ряд Тейлора характеризуется членами второго порядка малости.

Основными параметрами импеданса являются положения экстремумов амплитудно-частотной характеристики, которые соответствуют формантным частотам. В табл. 2.3 проведено сравнение положения максимумов входного импеданса речеобразующего тракта, рассчитанных на основе соотношения (2.4), со значениями формант для различных звуков. Наиболее существенные расхождения имеют место для высших формант. Такие же расхождения наблюдаются для моделей Келли-Локбаума [105, 116, 125] и др..

Таблица 2.3 Расчетные значения положения модуля входного импеданса

Частота (Гц)	F_1		F_2		F_3		F_4		F_5	
Звук	эксп	расчет	Эксп	Расчет	эксп	Расчет	эксп	Расчет	Эксп	расчет
А	700	700	1080	1100	2600	2700	3550	4100	4000	5050
О	535	500	780	980	2500	2600	3500	4200	3800	4800
У	300	300	625	810	2500	2700	3400	3550	3950	4650
И	240	240	2250	2500	3200	3300	3570		4400	4350
Ы	300	380	1480	1400	2330	2380	3200	3900	4200	4550
Э	440	440	1880	1480	2550	2550	3410	3800	4150	5000
С	483	480	1619	1600	2861	2500	4030	4200	4406	5200
Ш	473	480	1440	1550	2102	2400	2530	2800	3160	4400
Ф	340		1200		2110		2694		3873	
Х	544		1460		2035		2915		3700	

Проведенное моделирование показывает, что в основу модели входного импеданса речеобразующего тракта можно положить аналог электрической цепи с сосредоточенными параметрами. Более того, этот аналог может состоять из конечного, не очень большого (в отличие от модели LEA [125]), числа элементов, количество которых будет определяться формантной структурой речевого сигнала.

2.2 Аппроксимация входного импеданса речеобразующего тракта

Основная задача аппроксимации состоит в том, чтобы получить аналитические соотношения для входного импеданса речеобразующего тракта, позволяющие провести моделирование процессов образования звучной речи. Естественно, что при этом необходимо сохранить основные свойства входного импеданса голосообразующего тракта.

Задаче аппроксимации коэффициента передачи речеобразующего тракта уделялось в свое время большое внимание, и было предложено ряд моделей [108, 114, 115, 116, 117, 125, 127] той или иной степени сложности. Однако все

эти модели были ориентированы на получение коэффициента передачи в схеме речеобразования, предложенной Фантом [125].

По результатам анализа, проведенного в предыдущем разделе, было показано, что при аппроксимации в первом приближении можно ограничиться учетом частот первых пяти формант. Тем более что достоверных данных по частотам более высоких формант практически нет, а уровень речевого сигнала для более высоких частот очень низок (ослабление интенсивности речевого сигнала на частотах выше 1000 Гц в среднем составляют 6 Дб/окт на октаву).

Анализ резонансных свойств модели входного импеданса речеобразующего тракта показывает, что он приближается по своим характеристикам к реактивному двухполоснику. В этом случае задача аппроксимации заданного импеданса сводится к синтезу соответствующего реактивного двухполосника, имеющего входной импеданс вида:

$$Z(s) = A \frac{s \cdot \prod_{k=1}^m (s^2 + (\omega_k^0)^2)}{\prod_{i=1}^n (s^2 + (\omega_i^P)^2)} \quad (2.5)$$

где ω_k^0 и ω_i^P – нули и полюса входного импеданса, соответственно;

A – масштабный коэффициент.

Естественно использовать в качестве ω_i^P значения частот формант, полученных экспериментально (см. табл. 2.3). Положение нулей ω_k^0 на оси частот трудно поддается экспериментальному измерению и практически неизвестно. Поэтому за основу значений ω_i^P будем использовать их положения на оси частот, полученные для моделей предыдущего раздела. В табл. 2.4 приведены значения частот, соответствующие нулям входного импеданса, полученных на основании соотношения (2.4).

Таблица 2.4. Нули передаточной функции

Частота, $\Gamma_{\text{ц}}$	Нули F_1		Нули F_2		Нули F_3		Нули F_4	
Звук	расч.	скорр.	расч.	скорр.	расч.	скорр.	расч.	скорр.
А	790	786	2355	2272	2875	2720	4355	3856
О	750	663	2185	2059	2750	2594	4425	3745
У	690	549	1690	1498	2930	2744	3780	3672
И	770	711	2440	2180	3100	2926	4140	4261
Ы	990	1006	1750	1784	3110	2748	4160	3929
Э	780	911	2010	2212	2999	2859	4270	3880
С	690	696	1990	2157	3160	3315	4440	4266
Ш	875	830	1970	1767	2500	2209	3310	2799
Ф	650		1820		2860		4110	
Х	720		2005		3200		4640	

Так как расчетные значения формант не совпадают с экспериментальными, то значения частот F_k^0 были скорректированными на основе соотношений:

$$F_{iann}^0 = \frac{\alpha \cdot F_{i+1\varnothing}^P + F_{i\varnothing}^P}{\alpha + 1}, \quad (2.6)$$

где F_{iann}^0 - значения частоты i -го нуля при аппроксимации с использованием экспериментальных данных; $F_{\varnothing i}^P$ - частота i - й форманты; α - коэффициент коррекции;

$$\alpha = \alpha_{ann} = \alpha_{\varnothing} = \frac{F_{iann}^0 - F_{i\varnothing}^P}{F_{i+1\varnothing}^P - F_{iann}^0} = \frac{F_{ip}^0 - F_{ip}^P}{F_{i+1p}^P - F_{ip}^0}, \quad (2.7)$$

где α_{ann} и α_{\varnothing} - аппроксимирующие и расчетные значения коэффициента коррекции;

$F_{i\varnothing}^0$ и $F_{i\varnothing}^P$ - расчетные значения частот нулей и полюсов входного импеданса.

В табл. 2.4 приведены, кроме расчетных, скорректированные значения положения нулей.

Следующим этапом аппроксимации является выбор топологии соответствующего электрического аналога на сосредоточенных параметрах. Можно считать [125, 127], что при поперечных размерах тракта, меньших по сравнению с длиной волны, распределение скоростей частиц на поверхности

излучателя является однородным. Дифференциальная характеристика (скорости) обладает свойствами индуктивности, что позволяет определять нагрузку как индуктивную.

Исходя из данных предыдущего моделирования и результатов исследований, приведенных в работах [5, 6, 7, 8], за основу можно выбрать электрический аналог в виде цепной схемы, приведенной на рис. 2.3, для которой известны алгоритмы расчета параметров элементов цепи, например метод Кауэра [41].

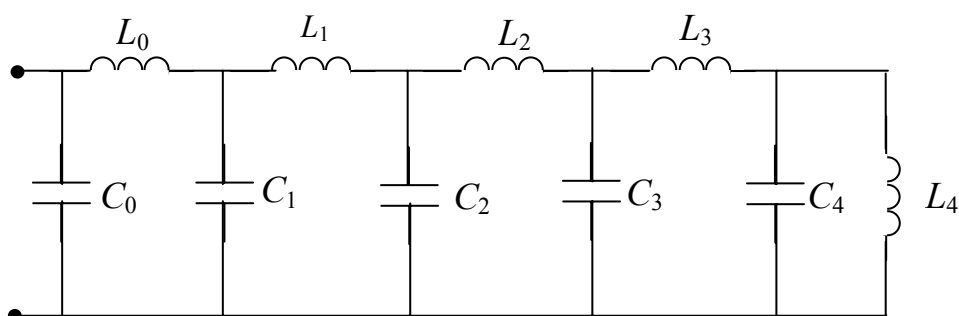


Рисунок 2.3 Участок электрического аналога речеобразующего тракта

Однако значения индуктивностей L_i , емкостей C_i при таком алгоритме расчета будут всегда нормированы так, что $C_0=1$. Поэтому необходимо эти значения скорректировать с учетом реальных объемов речеобразующего тракта для разных звуков. Будем исходить из предположения, что суммарная емкость электрического аналога

$$C = \sum_{i=0}^4 C_i$$

должна быть пропорциональна общему объему речеобразующего тракта от точки x_0 до губ. В табл. 2.5 приведены значения объемов речеобразующего тракта $V_{\text{рот}}$ и объемов гортанного желудочка $V_{\text{м}}$ для различных звуков, а в табл. 2.6 приведены расчетные значения емкостей и индуктивностей при значениях коэффициента $A=1$. В табл. 2.5 приведены также оценки емкости электрического аналога гортанного желудочка, которая определялась также пропорционально его объем и объем речеобразующего тракта.

Таблица 2.5

Звук	А	О	У	И	Э	Ы
$V_M, \text{см}^3$	4,73	4,73	5,31	2,57	4,73	6,38
$V_{\text{РОТ}}, \text{см}^3$	71,52	63,8	59,66	65,49	67,04	56,43
$C_M \cdot 10^{-4}, \Phi$	2,45	2,45	2,76	1,33	2,47	3,31

Таблица 2.6.

Звук	Первое звено		Второе звено		Третье звено		Четвертое звено		Выход схемы	
	L_0	C_0	L_1	C_1	L_2	C_2	L_3	C_3	L_4	C_4
	нГн	Ф	нГн	Ф	нГн	Ф	нГн	Ф	нГн	Ф
А	2.773	1	3.116	2.252	3.399	1.868	1.386	2.459	1.075	29.55
О	3.976	0.759	5.006	1.73	9.071	1.141	3.925	0.816	2.178	28.68
У	1.622	1.688	2.006	3.897	5.439	2.217	6.297	0.905	7.843	22.26
И	0.445	5.11	0.390	12.41	0.847	11.1	6.352	0.522	6.058	0.771
Ы	0.803	3.21	0.816	6.06	1.004	7.476	1.928	7.142	7.732	4.865
Э	0.639	4.116	0.615	9	0.76	9.455	1.666	6.849	1.993	5.373

В общем случае в схему следует включить потери. Одно звено с потерями цепной схемы (рис. 2.3) представлено на рис. 2.4.

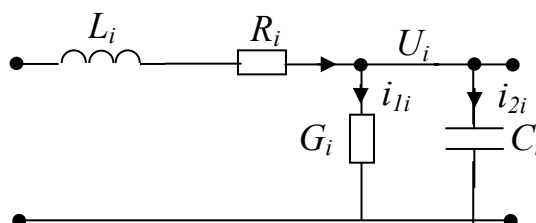


Рисунок 2.4 Схема электрического аналога одного звена с учетом потерь

Предварительную оценку характера потерь в одном сечении речеобразующего тракта можно сделать по данным табл. 2.1, из соотношений, приведенных в ней, следует, что потери зависят от частоты. На рис. 2.5 представлены графики зависимостей сопротивления и проводимости от частоты для двух крайних значений сечения речеобразующего тракта. Как следует из этих графиков, сопротивления и проводимости существенно зависят от площади поперечного сечения речеобразующего тракта. Однако наибольший интерес представляют постоянные времени:

$$\tau_{1i} = \frac{L_i}{R_i} \quad \text{и} \quad \tau_{2i} = \frac{C_i}{G_i}. \quad (2.8)$$

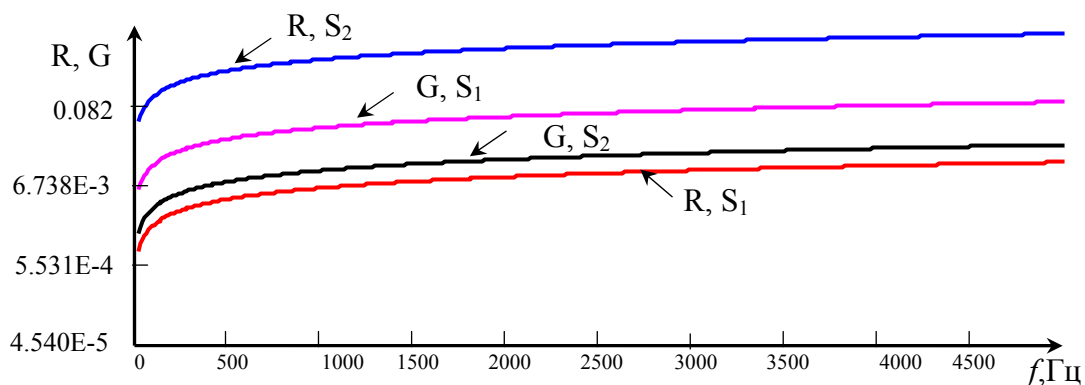


Рисунок 2.5 Зависимость сопротивления и проводимости от частоты, секции речеобразующего тракта для $S_1=9,7 \text{ см}^2$ и $S_2=0,7 \text{ см}^2$

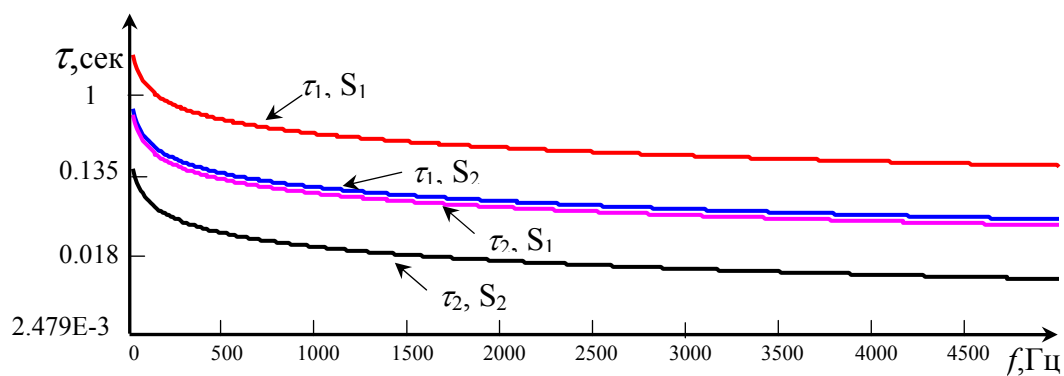


Рисунок 2.6 Зависимость постоянных времени τ_1 и τ_2 от частоты, секции речеобразующего тракта для $S_1=9,7 \text{ см}^2$ и $S_2=0,7 \text{ см}^2$

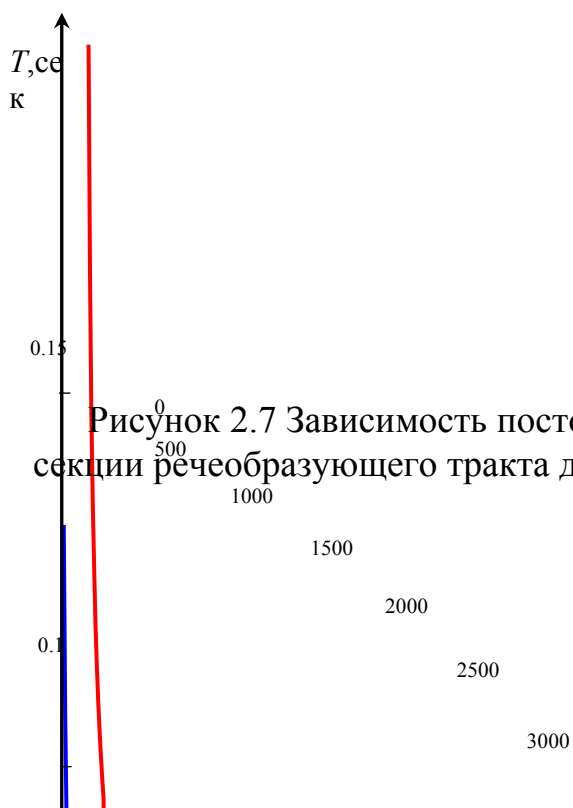


Рисунок 2.7 Зависимость постоянной времени от частоты, секции речеобразующего тракта для $S_1=9,7 \text{ см}^2$ и $S_2=0,7 \text{ см}^2$

На рис. 2.6 приведены расчетные зависимости постоянных времени τ_{1i} и τ_{2i} от частоты для этих же сечений речеобразующего тракта, которые также существенно зависят от площади поперечного сечения. Эквивалентная постоянная времени электрического аналога определяется соотношением

$$\tau_i = \frac{\tau_{1i} \cdot \tau_{2i}}{\tau_{1i} + \tau_{2i}}. \quad (2.9)$$

График зависимости эквивалентной постоянной времени для двух крайних сечений речеобразующего тракта приведен на рис. 2.7, из которого следует, что она существенно меньше зависит от площади поперечного сечения. Особенно в области частот выше 500 Гц.

Ясно, что учесть частотную зависимость эквивалентной постоянной времени в модели речеобразования достаточно сложно. Поэтому на первых этапах будем предполагать, что постоянная времени (2.9) не зависит от частоты и будет определять R и G для звена, представленного на рис. 2.4, в виде

$$R = b_1 L \text{ и } G = b_2 C, \quad (2.10)$$

$$\text{где } b_1 = \frac{1}{\tau_1}; \quad b_2 = \frac{1}{\tau_2}.$$

При этом будем учитывать, что

$$\tau_{3B} = \frac{1}{b_1 + b_2}. \quad (2.11)$$

Оценка значений R и C , а также τ будет проведена в третьей главе с учетом особенностей возбуждения устойчивых колебаний голосовых складок.

На рис. 2.8 представлены характеристики входного импеданса, полученные на основании соотношения (2.5) с учетом потерь (2.10), при различных значениях коэффициентов b_1, b_2 .

Сопоставляя входные импедансы речеобразующего тракта для модели (2.5) с сосредоточенными параметрами с моделью (2.4) с распределенными, можно отметить, что они эквивалентны. Следовательно, в дальнейшем можно пользоваться моделью входного импеданса речеобразующего тракта с сосредоточенными параметрами. Это позволяет проводить анализ процессов

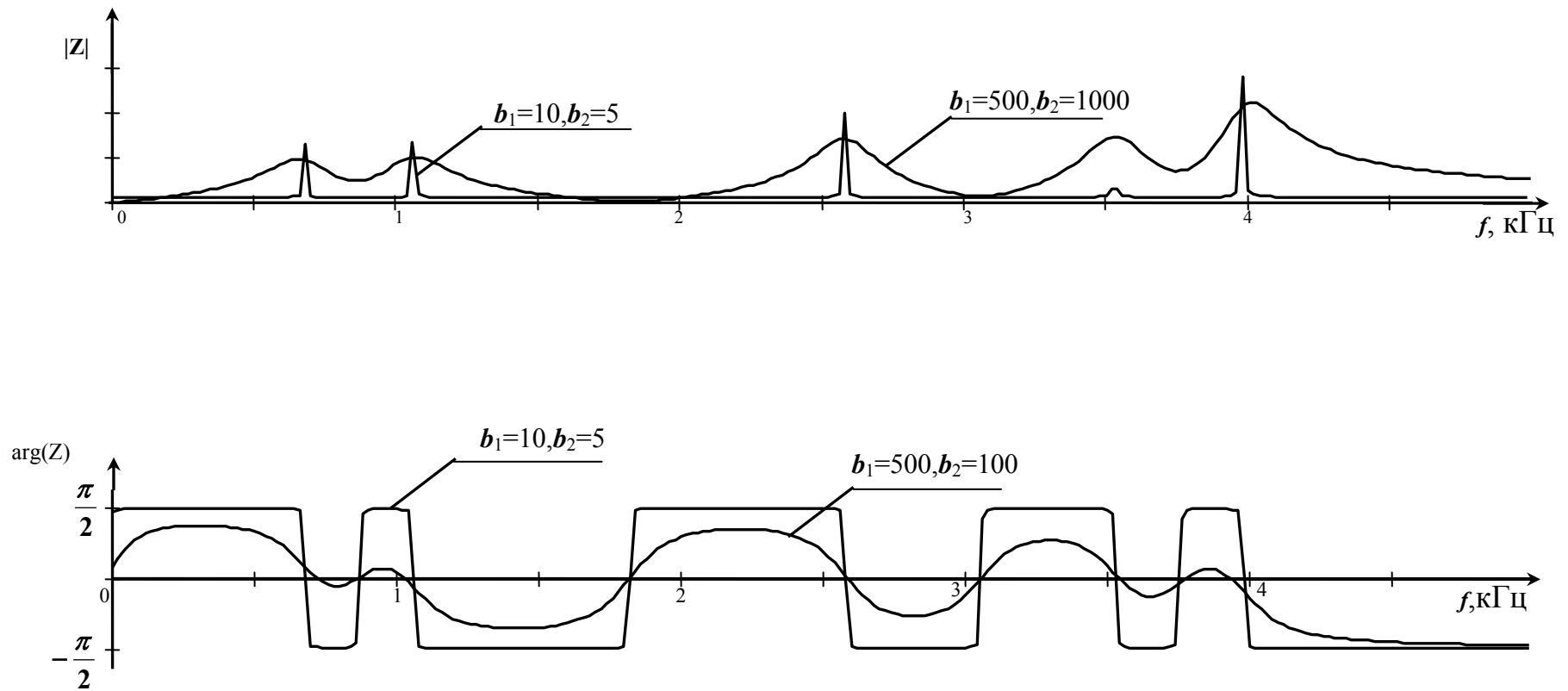


Рисунок 2.8 Амплитудно-частотная и фазо-частотные характеристики входного импеданса при разных значениях коэффициентов b_1 , b_2 для звука «А».

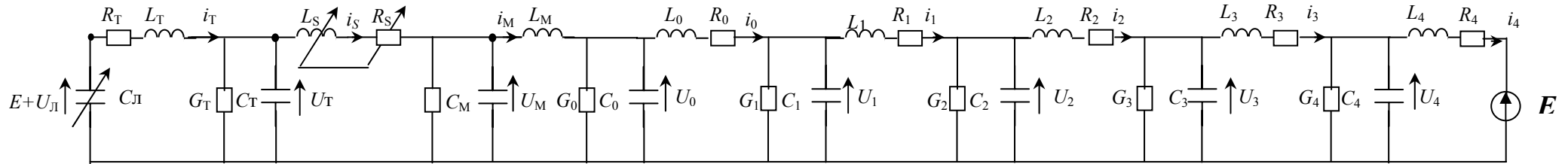


Рисунок 2.9 Схема электрического аналога голосообразующего тракта для нормальной речи. В этой модели C_L – емкость, представляющая эквивалентный объем легких; L_T , R_T , и C_T – параметры электрического эквивалента трахеи; L_S и R_S – параметры электрического эквивалента голосовой щели; E – эквивалент атмосферного давления

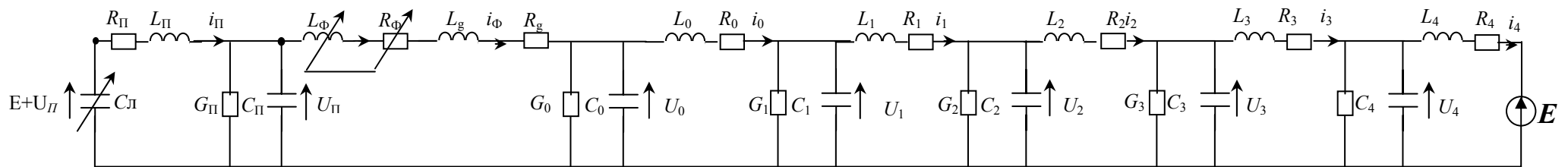


Рисунок 2.10 Схема электрического аналога голосообразующего тракта для пищеводной речи. Параметры воздушного пузыря и отрезка пищевода до сфинктера представлены элементами: C_B , R_{II} , L_{II} , G_{II} и C_{II}

голосообразования звучной речи на основе обычных дифференциальных уравнений.

2.3 Модель фонации для нормальной речи

Полученная модель входного импеданса речеобразующего тракта позволяет построить электроакустический аналог общей модели образования звучной речи, т.е. модель фонации.

Общая модель фонации представлена на рис. 2.9. В этой модели C_l – емкость, представляющая эквивалентный объем легких; L_T , R_T ; и C_T – параметры электрического эквивалента трахеи; L_C и R_C – параметры электрического эквивалента голосовой щели; E – эквивалент атмосферного давления.

Необходимо отметить, что подобные модели легких, трахеи и голосовой щели были предложены и исследованы в работах [10,12,13,14,15,18,59,141,142,143], где было показано, что они достаточно полно отражают основные свойства процессов образования звучной речи. Однако в подобных моделях других авторов [1,30,65,66,67,68,73,74,75,76] не учитывалось влияние разности давлений в трахее и глоточном желудочке на движение голосовых складок, а также воздушного потока через голосовые складки на разность давлений. Это не позволило получить качественный синтез речи.

Движение голосовых складок на этапе раскрытия в простейшем случае (одномерная модель) можно описать уравнением [117,127]

$$\frac{d^2x}{dt^2} + \delta \frac{dx}{dt} + \omega_c^2 x = F(t), \quad (2.12)$$

где x – среднее расстояние между голосовыми складками при их раскрытии; δ – потери при движении складок; ω_c – собственная резонансная частота голосовых складок.

Обычно [127] считается, что длина голосовой щели l постоянна. Следовательно, площадь раскрытия голосовых складок, приблизительно определяемая как

$$S = l \cdot x, \quad (2.13)$$

оказывается пропорциональна x . Таким образом, можно считать, что [190]

$$L_c = \frac{\alpha}{S}, \quad R_c = \frac{\beta}{S} \quad (2.14)$$

где α и β – коэффициенты пропорциональности.

Сила $F(t)$, действующая на голосовые складки, с учетом сил Бернулли, которая пропорциональна квадрату от скорости воздушного потока, может быть представлена соотношением [117]

$$F(t) = \frac{U_T - U_M}{a} - \frac{\gamma}{a} \cdot i_c^2, \quad (2.15)$$

где U_T, U_M – напряжения на выходе трахеи и в гортанном желудочке; a, γ – коэффициенты пропорциональности; i_c – ток через раскрытые голосовые складки. Очевидно, что при закрытии голосовых складок требуется некоторое время, в течение которого рассеивается кинетическая энергия, приобретенная голосовыми складками в процессе их раскрытия. Этот процесс можно смоделировать уравнениями, подобным уравнению (2.12), но с другими параметрами δ_l и ω_{cl} .

Начальный момент раскрытия голосовых складок может происходить по двум причинам. Во-первых, когда сила $F(t)$ превышает некоторое пороговое значение P , так называется миоэластической подход к модели образования звучной речи, т.е. при условии:

$$F(t) > P. \quad (2.16)$$

Во-вторых, под действием нервного возбуждения (нейрохроностическая теория). В этом случае будем считать, что сила, действующая на голосовые складки, имеет вид

$$F_1(t) = F_H(t) + B \cdot F(t), \quad (2.17)$$

где $F_H(t)$ – сила, действующая на голосовые складки, определяемая нервным возбуждением;

$B \leq 1$ – коэффициент, учитывающий влияние на движение голосовых складок разности давлений в трахее и глоточном желудочке.

В первом приближении движущая сила на голосовые складки за счет нервного возбуждения может быть представлена периодической последовательностью дельта импульсов, следующих с частотой основного тона.

Все составляющие реальной системы речеобразования, т.е. легкие, трахея, речеобразующий тракт, после вдоха в исходный момент времени находятся под атмосферным давлением, эквивалент которого в модели напряжение E . Голосовые складки сомкнуты, т.е. легкие и трахея изолированы от речеобразующего тракта, Все воздушные потоки практически равны нулю. Начало фонации начинается с момента уменьшения объема легких, (эквивалент емкости C_L в модели), что приводит к появлению избыточного давления по отношению к атмосферному.

В рассматриваемой модели за счет того, что потери, определяемые сопротивлениями R_i и проводимостями G_i , не равны нулю на нулевой частоте, возможны токи утечки. Эти токи приведут к снижению напряжения в различных точках схемы. Однако это всегда может быть учтено при проведении соответствующего анализа. В любом случае, процесс образования избыточного давления может быть описан уравнением [13, 141]:

$$\frac{dU_L}{dt} = -\frac{i_T}{C_0 - C} + \frac{E + U_L}{C_0 - C} \cdot \frac{dC}{dt}, \quad (2.18)$$

где U_L – избыточное напряжение;

C_0 – исходное значение емкости C_L ;

C – изменение значения емкости C_L ;

i_T – ток трахеи.

В данной схеме выполняется условие:

$$C_L = C_0 - C \gg C_T + C_M + \sum_{i=0}^4 C_i. \quad (2.19)$$

Уравнения, определяющие токи и напряжения в трахее, имеют вид

$$\begin{aligned}\frac{di_T}{dt} &= \frac{U_{\text{Л}} - U_T - i_T \cdot R_T}{L_T}; \\ \frac{dU_T}{dt} &= \frac{i_T - i_C - U_T \cdot G_T - G_T \cdot E}{C_T}.\end{aligned}\quad (2.20)$$

Ток, протекающий через разомкнутые голосовые складки, определяется соотношением:

$$\frac{di_C}{dt} = \frac{U_T - U_M}{L_C} - \left(\frac{R_C}{L_C} + \frac{1}{L_C} \cdot \frac{dL_C}{dt} \right) \cdot i_C - \frac{\lambda}{L_C} \cdot i_C^2, \quad (2.21)$$

где λ – коэффициент пропорциональности, учитывающий снижение давления за счет сил Бернулли.

Учитывая соотношение (2.14), выражение (2.21) принимает вид:

$$\frac{di_C}{dt} = \frac{U_T - U_M}{\alpha} S - \left(\frac{\beta}{\alpha} - \frac{x_1}{S} \right) \cdot i_C - \frac{\lambda}{\alpha} \cdot S \cdot i_C^2, \quad (2.22)$$

где $S = \frac{dS}{dt}$.

Естественно, что при сомкнутых голосовых складках, т.е. при $L_C \rightarrow \infty$, $R_C \rightarrow \infty$ и при $x = 0$, i_C равно нулю.

Величина

$$\frac{\alpha}{\beta} = \frac{L_C}{R_C} = \tau_C, \quad (2.23)$$

характеризует постоянную времени раскрытой голосовой щели.

Уравнения, описывающие токи и напряжения в i -ом звене речеобразующего тракта, имеют вид

$$\begin{aligned}\frac{di_i}{dt} &= \frac{U_i - U_{i+1} - R_i i_i}{L_i}, \\ \frac{dU_i}{dt} &= \frac{i_{i-1} - i_i - U_i G_i - E \cdot G_i}{L_i}.\end{aligned}\quad (2.24)$$

Аналогичные уравнения имеют место и для тока и напряжения глоточного желудочка.

Полученные соотношения (2.12), (2.18), (2.20), (2.22), (2.24) образуют систему дифференциальных уравнений. Полностью она представлена в табл. 2.7. Данная система уравнений позволяет провести анализ процессов

образования звучной речи для двух случаев возбуждения голосовых складок, описываемых соотношениями (2.16) и (2.17).

Таким образом, построены модели фонации для нормальной речи, учитывающие две противоречивые теории возникновения колебания голосовых складок: миоэластическую и нейрохроностическую.

2.4 Модель фонации для пищеводной речи

Модель фонации для пищеводной речи может быть получена на основе модели голосообразования для нормальной речи. Как уже отмечалось в первой главе, анатомия речеобразующего тракта в обоих случаях остается неизменной, что отражено на рис. 2.10, где представлена модель голосообразования для пищеводной речи.

Основные изменения касаются источника избыточного давления, которое формируется за счет воздушного пузыря в пищеводе. Параметры воздушного пузыря и отрезка пищевода до псевдоголосовой щели представлены элементами: C_B , R_{II} , L_{II} , G_{II} и C_{II} .

Можно предположить, что соотношение (2.19) для емкости C_B не выполняется, т.е. емкость воздушного пузыря сопоставима с емкостью речеобразующего тракта:

$$C_B = C_0 - C \geq C_{II} + \sum_{i=0}^4 C_i. \quad (2.25)$$

А это значит, что следует ожидать существенного влияния источника избыточного давления на процесс фонации.

По аналогии с моделью образования нормальной речи, процессы происходящие в воздушном пузыре, можно описать уравнениями (2.18) и (2.20), производя соответствующую замену переменных.

Будем считать, что движение псевдоголосовой щели описывается также уравнением типа (2.12) и что вынужденная сила также может быть представлена в форме (2.15), (2.16) и (2.17). Кроме того, параметры L_Φ и R_Φ также подчиняются закономерностям (2.14).

Однако уравнение, описывающее ток через псевдоголосовую щель, будет иметь вид:

$$\frac{di_{\phi}}{dt} = \frac{U_{\Pi} - U_O}{L_{\phi} + L_g} - \left(\frac{R_{\phi} + R_g}{L_{\phi} + L_g} + \frac{1}{L_{\phi} + L_g} \cdot \frac{dL_{\phi}}{dt} \right) \cdot i_{\phi} - \frac{\lambda}{L_{\phi} + L_g} \cdot i_{\phi}^2. \quad (2.26)$$

Тогда с учетом (2.14), выражение (2.26) приобретает вид:

$$\frac{di_{\phi}}{dt} = \frac{U_{\Pi} - U_O}{\alpha + L_g \cdot S} \cdot S - \left(\frac{\beta + R_g \cdot S}{\alpha + L_g \cdot S} - \frac{x1}{S \cdot (\alpha + L_g \cdot S)} \right) \cdot i_{\phi} - \frac{\lambda \cdot S}{\alpha + L_g \cdot S} \cdot i_{\phi}^2, \quad (2.27)$$

Из соотношения (2.24) следует, что постоянная времени для тока i_{ϕ} при раскрытии псевдоголосовой щели теперь будет функцией S , т.е.

$$\tau_n = \frac{\alpha + L_g \cdot x}{\beta + R_g \cdot x}, \quad (2.28)$$

и только при малых значениях S будем приближаться к постоянной времени раскрытой псевдоголосовой щели. Естественно, что это может привести к резким выбросам напряжения на входе речеобразующего тракта в моменты схлопывания сфинктера. Поэтому следует ожидать резкое повышение уровня высокочастотных гармоник в пищеводной речи.

Уравнения, описывающие токи и напряжения в i -м звене речеобразующего тракта, остаются без изменений, т.е. совпадают с уравнениями (2.24) для процессов, описывающих образования нормальной речи.

Полученные соотношения для модели образования нормальной речи, с учетом выражения (2.27), образуют систему дифференциальных уравнений для модели образования пищеводной речи. Полностью она приведена в табл. 2.8.

Таким образом, построена модель фонации для пищеводной речи. Качественный анализ постоянных времени развития процессов формирования тока возбуждения, т.е. тока через псевдоголосовую щель, показывает, что пищеводная речь по сравнению с нормальной должна иметь повышенный уровень высокочастотных гармоник. Кроме того, ограниченность объема воздушного пузыря должна привести к существенной неустойчивости процесса фонации, если процесс раскрытия псевдоголосовой щели определяется соотношением (2.17) (миоэластический подход к модели образования звуковой речи).

Таблица 2.7 Система уравнений для образования нормальной речи $U_L = E - \Delta U_L$
 $C = C_0 - C$

Действующие токи и напряжения	Действующие токи и напряжения в модели	Обозначения в модели
$\frac{dU_L}{dt} = \frac{U_L}{C} \frac{dC}{dt} - \frac{i_T}{C}$	$\frac{dy_0}{dt} = \frac{-y_0}{C} \frac{dC}{dt} - \frac{y_1}{C}$	$U_L = y_0$
$\frac{di_T}{dt} = \frac{U_L - U_T - i_T \cdot R_T}{L_T}$	$\frac{dy_1}{dt} = \frac{y_0 - y_2 - y_1 \cdot R_T}{L_T}$	$i_T = y_1$
$\frac{dU_T}{dt} = \frac{1}{C_T} (i_T - i_C - G_T U_T)$	$\frac{dy_2}{dt} = \frac{y_1 - y_5 - y_2 \cdot G_T}{C_T}$	$U_T = y_1$
$\frac{dx}{dt} = x1$	$\frac{dy_3}{dt} = y_4$	$x = y_3$
$\frac{dx1}{dt} = F(t) - \delta \cdot x1 - \omega_c^2 \cdot x$	$\frac{dy_4}{dt} = \frac{y_2 - y_6}{a} - \gamma \cdot y_5^2 - \delta \cdot y_4 - \omega_c^2 \cdot y_3$	$x_1 = y_4$
$\frac{di_C}{dt} = \frac{U_T - U_M}{\alpha} x - \left(\frac{\beta}{\alpha} - \frac{x1}{x} \right) \cdot i_C - \frac{\lambda}{\alpha} x \cdot i_C^2$	$\frac{dy_5}{dt} = \frac{y_2 - y_6}{\alpha} y_3 - \left(\frac{\beta}{\alpha} - \frac{y_4}{y_3} \right) \cdot y_5 - \frac{\lambda}{\alpha} y_5 \cdot y_3^2$	$i_C = y_5$
$\frac{dU_M}{dt} = \frac{i_C - i_M - U_M \cdot G_M}{C_M}$	$\frac{dy_6}{dt} = \frac{y_5 - y_7 - y_6 \cdot G_M}{C_M}$	$U_M = y_6$
$\frac{di_M}{dt} = \frac{U_M - U_O - i_M \cdot R_M}{L_M}$	$\frac{dy_7}{dt} = \frac{y_6 - y_8 - y_7 \cdot R_M}{L_M}$	$i_M = y_7$
$\frac{dU_O}{dt} = \frac{i_M - i_O - U_O \cdot G_O}{C_O}$	$\frac{dy_8}{dt} = \frac{y_7 - y_9 - y_8 \cdot G_O}{C_O}$	$U_O = y_8$
$\frac{di_O}{dt} = \frac{U_O - U_1 - i_O \cdot R_O}{L_O}$	$\frac{di_9}{dt} = \frac{y_8 - y_{10} - y_9 \cdot R_O}{L_O}$	$i_O = y_9$
$\frac{dU_1}{dt} = \frac{i_O - i_1 - U_1 \cdot G_1}{C_1}$	$\frac{dy_{10}}{dt} = \frac{y_9 - y_{11} - y_{10} \cdot G_1}{C_1}$	$U_1 = y_{10}$
$\frac{di_1}{dt} = \frac{U_1 - U_2 - i_1 \cdot R_1}{L_1}$	$\frac{dy_{11}}{dt} = \frac{y_{10} - y_{12} - y_{11} \cdot R_1}{L_1}$	$i_1 = y_{11}$
$\frac{dU_2}{dt} = \frac{i_1 - i_2 - U_2 \cdot G_2}{C_2}$	$\frac{dy_{12}}{dt} = \frac{y_{11} - y_{13} - y_{12} \cdot G_2}{C_2}$	$U_2 = y_{12}$
$\frac{di_2}{dt} = \frac{U_2 - U_3 - i_2 \cdot R_2}{L_2}$	$\frac{dy_{13}}{dt} = \frac{y_{12} - y_{14} - y_{13} \cdot R_2}{L_2}$	$i_2 = y_{13}$
$\frac{dU_3}{dt} = \frac{i_2 - i_3 - U_3 \cdot G_3}{C_3}$	$\frac{dy_{14}}{dt} = \frac{y_{13} - y_{15} - y_{14} \cdot G_3}{C_3}$	$U_3 = y_{14}$
$\frac{di_3}{dt} = \frac{U_3 - U_4 - i_3 \cdot R_3}{L_3}$	$\frac{dy_{15}}{dt} = \frac{y_{14} - y_{16} - y_{15} \cdot R_3}{L_3}$	$i_3 = y_{15}$
$\frac{dU_4}{dt} = \frac{i_3 - i_4 - U_4 \cdot G_4}{C_4}$	$\frac{dy_{16}}{dt} = \frac{y_{15} - y_{17} - y_{16} \cdot G_4}{C_4}$	$U_4 = y_{16}$
$\frac{di_4}{dt} = \frac{U_4 - E - i_4 \cdot R_4}{L_4}$	$\frac{dy_{17}}{dt} = \frac{y_{16} - E - y_{17} \cdot R_4}{L_4}$	$i_4 = y_{17}$

Таблица 2.8 Система уравнений для образования пищеводной речи

$$U_{\Pi Y} = E - \Delta U_{\Pi Y} \quad C = C_0 - C$$

Действующие токи и напряжения	Действующие токи и напряжения в модели	Обозначения в модели
$\frac{dU_{\Pi Y}}{dt} = \frac{U_{\Pi Y}}{C_{\Pi Y}} \frac{dC_{\Pi Y}}{dt} - \frac{i_{\Pi}}{C_{\Pi Y}}$	$\frac{dy_0}{dt} = \frac{y_0}{C_{\Pi Y}} \frac{dC_{\Pi Y}}{dt} - \frac{y_1}{C_{\Pi Y}}$	$U_{\Pi Y} = y_0$
$\frac{di_{\Pi}}{dt} = \frac{U_{\Pi Y} - U_{\Pi} - i_{\Pi} \cdot R_{\Pi}}{L_{\Pi}}$	$\frac{dy_1}{dt} = \frac{y_0 - y_2 - y_1 \cdot R_{\Pi}}{L_{\Pi}}$	$i_{\Pi} = y_1$
$\frac{dU_{\Pi}}{dt} = \frac{1}{C_{\Pi}} (i_{\Pi} - i_{\phi} - G_{\Pi} U_{\Pi})$	$\frac{dy_2}{dt} = \frac{y_1 - y_5 - y_2 \cdot G_{\Pi}}{C_{\Pi}}$	$U_{\Pi} = y_1$
$\frac{dx}{dt} = x1$	$\frac{dy_3}{dt} = y_4$	$x = y_3$
$\frac{dx1}{dt} = F(t) - \delta \cdot x1 - \omega_c^2 \cdot x$	$\frac{dy_4}{dt} = \frac{y_2 - y_6 - \gamma \cdot y_5^2 - \delta \cdot y_4 - \omega_c^2 \cdot y_3}{a}$	$x_1 = y_4$
$\frac{di_{\phi}}{dt} = \frac{U_{\Pi} - U_0}{\alpha - L_g \cdot x} x - \left(\frac{\beta + R_g x}{\alpha + L_g x} - \frac{\alpha \cdot x1}{x} \right) \cdot i_{\phi} - \frac{\lambda}{\alpha + L_g x} x$	$\frac{dy_5}{dt} = \frac{y_2 - y_6}{\alpha + L_g y_3} y_3 - \left(\frac{\beta + R_g y_3}{\alpha + L_g y_3} - \frac{\alpha \cdot y_4}{y_3 (\alpha + L_g y_3)} \right) \cdot y_5 - \frac{\lambda \cdot y_3}{\alpha + L_g y_3} \cdot y_5^2$	$i_{\phi} = y_5$
$\frac{dU_0}{dt} = \frac{i_{\phi} - i_0 - U_0 \cdot G_0}{C_0}$	$\frac{dy_6}{dt} = \frac{y_5 - y_7 - y_6 \cdot G_0}{C_0}$	$U_0 = y_6$
$\frac{di_0}{dt} = \frac{U_0 - U_1 - i_0 \cdot R_0}{L_0}$	$\frac{dy_7}{dt} = \frac{y_6 - y_8 - y_7 \cdot R_0}{L_0}$	$i_0 = y_7$
$\frac{dU_1}{dt} = \frac{i_0 - i_1 - U_1 \cdot G_1}{C_1}$	$\frac{dy_8}{dt} = \frac{y_7 - y_9 - y_8 \cdot G_1}{C_1}$	$U_1 = y_8$
$\frac{di_1}{dt} = \frac{U_1 - U_2 - i_1 \cdot R_1}{L_1}$	$\frac{dy_9}{dt} = \frac{y_8 - y_{10} - y_9 \cdot R_1}{L_1}$	$i_1 = y_9$
$\frac{dU_2}{dt} = \frac{i_1 - i_2 - U_2 \cdot G_2}{C_2}$	$\frac{dy_{10}}{dt} = \frac{y_9 - y_{11} - y_{10} \cdot G_2}{C_2}$	$U_2 = y_{10}$
$\frac{di_2}{dt} = \frac{U_2 - U_3 - i_2 \cdot R_2}{L_2}$	$\frac{dy_{11}}{dt} = \frac{y_{10} - y_{12} - y_{11} \cdot R_2}{L_2}$	$i_2 = y_{11}$
$\frac{dU_3}{dt} = \frac{i_2 - i_3 - U_3 \cdot G_3}{C_3}$	$\frac{dy_{12}}{dt} = \frac{y_{11} - y_{13} - y_{12} \cdot G_3}{C_3}$	$U_3 = y_{12}$
$\frac{di_3}{dt} = \frac{U_3 - U_4 - i_3 \cdot R_3}{L_3}$	$\frac{dy_{13}}{dt} = \frac{y_{12} - y_{14} - y_{13} \cdot R_3}{L_3}$	$i_3 = y_{13}$
$\frac{dU_4}{dt} = \frac{i_3 - i_4 - U_4 \cdot G_4}{C_4}$	$\frac{dy_{14}}{dt} = \frac{y_{13} - y_{15} - y_{14} \cdot G_4}{C_4}$	$U_4 = y_{14}$
$\frac{di_4}{dt} = \frac{U_4 - E - i_4 \cdot R_4}{L_4}$	$\frac{dy_{15}}{dt} = \frac{y_{14} - E - y_{15} \cdot R_4}{L_4}$	$i_4 = y_{15}$

Выводы по главе:

Показано, что в качестве модели речеобразующего тракта может использоваться электрический аналог на сосредоточенных параметрах, это позволяет свести задачу к решению системы обыкновенных дифференциальных уравнений.

Предложен подход к оценке параметров электрического аналога речеобразующего тракта. Выбрана топология схемы электрического аналога и проведена оценка значений элементов этой схемы.

Построены модели голосообразования для нормальной и пищеводной речи, которые отличаются от известных тем, что они позволяют учитывать влияние перепада подскладочного и надскладочного давлений на процесс фонации.

Построенные модели позволяют провести исследование механизмов голосообразования на основании двух гипотез: милоэластической и нейрорхностической, а также проанализировать влияние сил Бернулли. Выявлены качественные различия в процессах голосообразования нормальной и пищеводной звуковой речи. Показано, что следует ожидать существенного повышения уровня высокочастотных компонентов в пищеводной речи, а также существенной неустойчивости частоты основного тона.

ГЛАВА 3. ИССЛЕДОВАНИЕ МОДЕЛЕЙ ОБРАЗОВАНИЯ ЗВУЧНОЙ РЕЧИ

Во второй главе построены электрические аналоги моделей образования звучной нормальной и пищеводной речи. Эти модели описываются системами нелинейных дифференциальных уравнений с параметрами, в общем случае зависящими от времени. Однако на стационарном участке генерации звука, при определенных условиях, их можно свести к уравнениям с постоянными коэффициентами, т.е. к автономным системам.

Для полученных систем уравнений не определены ряд параметров, поэтому на первом этапе необходимо провести оценку значений этих параметров, определяющих постоянные времени и потери в различных звеньях моделей. Ясно, что эти параметры могут и должны принимать значения в широком диапазоне. Это объясняется тем, что, во-первых, анатомия речеобразующих систем для различных людей существенно отличается друг от друга и, во-вторых, генерировать речевой сигнал способны, например, птицы, животные, имеющие весьма отличную анатомическую систему.

Естественно ожидать, что для полученной системы уравнений должны существовать такие диапазоны значений параметров, когда колебания в системе устойчивы. По-видимому, области устойчивых колебаний позволят уточнить параметры системы. Поэтому на втором этапе необходимо провести исследования поведения системы с целью определения диапазонов параметров, обеспечивающих появление устойчивых колебаний.

3.1 Оценка параметров моделей образования звучной речи

Точно определить параметры моделей, представленных на рис. 2.9 и рис. 2.10, в настоящее время практически невозможно. Это объясняется тем, что данные по сечениям, объемам, а также другими анатомическим и физическими характеристиками речеобразующей системы весьма приближены и во многом противоречивы [107,116,117,123,125,127], особенно для динамики

процессов образования речи. Существующие в настоящее время технические возможности не позволяют провести точные измерения параметров речеобразующей системы. Кроме того, данные по анатомии речеобразующей системы являются неполными, еще нет достоверных зависимостей, устанавливающих соотношения между объемами, сечениями и физиологическими параметрами стенок и емкостями, индуктивностями и потерями соответственно.

Однако приведенные оценки в работах [108,116,117,125,126], имеющих характер обзора, позволяют приближенно задать некоторые опорные значения параметров речеобразующей системы и затем провести анализ поведения моделей в некотором диапазоне их значений. Можно ожидать, что такой подход позволит уточнить параметры электроакустических моделей, которые будут, в некоторой степени, адекватны реальным системам речеобразования.

Наиболее просто определяются требования к формированию напряжения U_L , которое является аналогом избыточного давления. По данным, обобщенным в [115,125,127], фонация начинается при давлении в легких около 1300 мм вод. ст. Это позволяет оценить величину

$$P_{изб} \geq \frac{U_L}{E} \quad (3.1)$$

значением порядка 1,2-1,3. В дальнейшем будем считать, что для избыточного давления в воздушном пузыре при образовании пищеводной звучной речи справедливы те же самые соотношения.

Для того чтобы более точно установить соотношение между параметрами речеобразующей системы и ее электрическим аналогом, необходимо провести анализ размерности величин, которые устанавливают связь между параметрами этих систем. Естественно, что ряд параметров не требуют более детального анализа их размерностей, так как необходимое соответствие было установлено на этапе их синтеза. Это в первую очередь касается электрического аналога речеобразующего тракта.

В табл. 3.1 приведены основные параметры, их обозначения и размерности, которые устанавливают соответствие между системой речеобразования и ее электрическим аналогом. Как и ожидалось, большинство из них (шесть из семи) связаны с функционированием голосовых складок.

Таблица 3.1 Размерности величин

Наименование параметра	Обозначение	Размерность	Производная единица СИ
Площадь раскрытия голосовой щели	S	L^2	m^2
Постоянная времени раскрытия голосовых складок	$\tau_{rc} = \frac{1}{\delta}$	T	C
Масштабный коэффициент реактивных параметров речеобразующего тракта	A	$L^2 M T^4 I^2$	
Коэффициент соответствия между площадью голосовой щели и ее индуктивностью.	α	$L^4 M T^2 I^2$	$Гн м^2$
Коэффициент соответствия между площадью голосовой щели и ее сопротивлением.	β	$L^4 M T^3 I^2$	$Ом м^2$
Коэффициент соответствия между давлением на голосовые складки и напряжением.	a	$M T^{-1} I^1$	$\frac{В \cdot c^2}{м^2}$
Коэффициенты, учитывающие влияние сил Бернулли.	γ, λ	$L^2 M T^3 I^3$	$\frac{кг \cdot м^2}{с^3 \cdot А^3}$

Дальнейшую оценку параметров моделей речеобразования будем проводить, привязываясь к параметрам основной части речеобразующей системы – голосовым складкам или их эквивалентам. Основным параметром, характеризующим голосовые складки в процессе фонации, является площадь раскрытия голосовой щели.

По оценкам большинства авторов [115,116,125,127], наибольшая площадь раскрытия голосовой щели составляет

$$S_{\max} = (0,1 \div 0,4) \cdot 10^{-4} \text{ м}^2. \quad (3.2)$$

При этом верхняя оценка считается завышенной.

В работах [125,127] приводится оценка постоянной времени (2.23) воздушного потока раскрытой голосовой щели. Она составляет

$$\tau \approx 0,25 \cdot 10^{-3} \text{ с}. \quad (3.3)$$

Однако необходимо учитывать, что потери в голосовой щели имеют две составляющих. Одна определяется сопротивлением R_c , а другая коэффициентом λ . С учетом этого, следует ожидать, что для (2.23)

$$\tau_c > \tau \quad \text{или} \quad R_c = b_c L_c, \quad (3.4)$$

$$\text{где } b_c = \frac{1}{\tau_c} = \frac{\beta}{\alpha} < \frac{1}{\tau}.$$

Исходя из этого, будем считать, что b_c лежит в пределах

$$b_c = 200 \div 2000. \quad (3.5)$$

Собственная резонансная частота колебаний раскрытых голосовых складок, по-видимому, близка к частоте F_0 основного тона, т.е. лежит в диапазоне

$$\omega_c = 2 \cdot \pi \cdot (80 - 300) \text{ рад/с}. \quad (3.6)$$

Неопределенным является значение потерь δ , характеризующих декремент затухания. Однако, можно предположить [108,116,125,127], что для выражения (2.9) ослабление колебаний в e раз происходит на 3-5 периодах основного тона, т.е.

$$\delta \approx \frac{F_0}{3 \div 5} = \frac{1}{(3 \div 5) \cdot T_0}, \quad (3.7)$$

где T_0 – период частоты основного тона F_0 .

Постоянная времени одного Г-образного звена в модели рис. 2.6 речеобразующего тракта определяется соотношением

$$\tau_{3B} = \frac{1}{b_1 + b_2}. \quad (3.8)$$

Эта величина характеризует декремент затухания, т.е. ширину резонансов (формант) для речеобразующего тракта. По данным [108,127], где приводятся оценки полосы частот формант, а также по виду амплитудно-частотных характеристик для входного импеданса речеобразующего тракта, приведенные во второй главе, можно предположить, что средняя:

$$\tau_{3B} \approx \frac{10 - 100}{F_{CP}}, \quad (3.9)$$

где F_{CP} – средняя частота первой форманты.

Тогда

$$b_1 + b_2 \approx 5 \div 50. \quad (3.10)$$

Известно, что минимальные сечения речеобразующего тракта много больше, чем наибольшая площадь раскрытия голосовой щели. Учитывая также, что потери обратно пропорциональны площади сечения [127], соотношения (3.9) и (3.10) можно взять за основу.

Приведенные соотношения (3.3)-(3.10), с учетом параметра масштабирования A в выражении (2.5), а также табл. 2.5, позволяют установить взаимосвязь между параметрами голосовой щели и параметрами речеобразующей системы. По-видимому, все это справедливо и для параметров гортанного желудка.

По данным [116,117] можно предположить, что собственная, т.е. резонансная частота трахеи, близка или кратна частоте основного тона. Однако, следует ожидать, что индуктивность трахеи больше суммарной индуктивности речеобразующего тракта, т.е.:

$$L_T > \sum_{i=0}^4 L_i. \quad (3.11)$$

На первом этапе также будем считать, что потери в трахее близки к потерям в речеобразующем тракте.

Остается неясным вопрос с коэффициентами λ и γ , которые учитывают влияния сил Бернулли на воздушный поток через голосовую щель и движение голосовых складок.

Анализ уравнений (2.14), (2.16) и (2.23) позволяет сделать вывод, что влияние сил Бернулли на воздушный поток и движение голосовых складок одинаково. Это подтверждается и совпадением размерностей коэффициентов λ и γ . Поэтому в дальнейшем будем полагать, что

$$\lambda = \gamma. \quad (3.12)$$

Таким образом, проведена предварительная оценка значений параметров моделей образования звучной речи. Это позволяет ограничить возможные

диапазоны изменений основных параметров моделей при исследовании их поведения.

3.2 Параметры и роль гортанного желудочка в процессе образования звучной речи

Сечения речеобразующего тракта, как уже отмечалось, характеризуются резкими сужениями в точке x_0 (см. табл. 2.2). Это сужение отделяет объем гортанного желудочка от остальной части речеобразующего тракта. Размеры гортанного желудочка не велики, т.е. его можно рассматривать как некоторый аналог электрической цепи с сосредоточенными параметрами, что отражено на модели, приведенной на рис. 2.12.

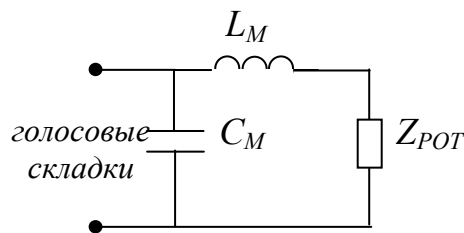


Рисунок 3.1. Электрический аналог гортанного желудочка

На рис. 3.1 приведена схема электрического аналога гортанного желудочка, которая включает емкость C_M , индуктивность L_M , а также входной импеданс $Z_{РОТ}$ речеобразующего тракта. В схему не включены элементы R_M и G_M в предположении, что потери невелики. В табл. 2.5 приведены оценки емкости C_M , но нет оценок для L_M . Следует ожидать, что индуктивность L_M обратно пропорциональна площади сечения речеобразующего тракта в точке x_0 . В табл. 3.1 приведены площади сечения для разных звуков в точке x_0 , на основе чего можно предположить, что значение данной индуктивности практически не зависит от произносимого звука.

Таблица 3.1 Расстояние до сужения

Звук	А	О	У	И	Э	Ы
$S (см^2)$	0,65	0,65	0,67	0,68	0,65	0,65

Оценить значение L_M можно по сравнению с суммарной индуктивностью всего речеобразующего тракта. В самом деле, входной импеданс речеобразующего тракта на низких частотах носит индуктивный характер и на частотах, близких к нулю, значение этой индуктивности равно

$$L = \sum_{i=0}^4 L_i. \quad (3.10)$$

Можно предположить, что

$$L_M \geq L. \quad (3.11)$$

Условие (3.11) можно обосновать тем, что площадь сечения речеобразующего тракта в точке x_0 , как правило, наименьшая. В табл. 3.2 приведены значения суммарной индуктивности и емкости речеобразующего тракта для модели, представленной на рис. 2.12.

Таблица 3.2 Параметры резонатора гортанного желудочка

Звук	А	О	У	И	Э	Ы
$L, \text{ нГн}$	11,75	24,16	23,25	14,09	5,67	12,28
C, Φ	37,129	33,127	30,967	34,012	34,79	29,3

На рис. 3.2 приведены зависимости частоты формант (резонансы входного импеданса голосообразующего тракта) и резонансы входного импеданса Z_M в зависимости от индуктивности L_M , на рис. 3.3 - модуль входного импеданса речеобразующего тракта и модуль Z_M при двух значениях индуктивности L_M , для звука «А». В приложении 5 приведены аналогичные графики для других звуков. Из представленных зависимостей следует, что с ростом индуктивности L_M резонансы Z_M начинают смещаться к положению формант. При этом уровень второго и третьего резонанса начинает падать и при больших значениях L_M может остаться только один экстремум в $|Z_M|$.

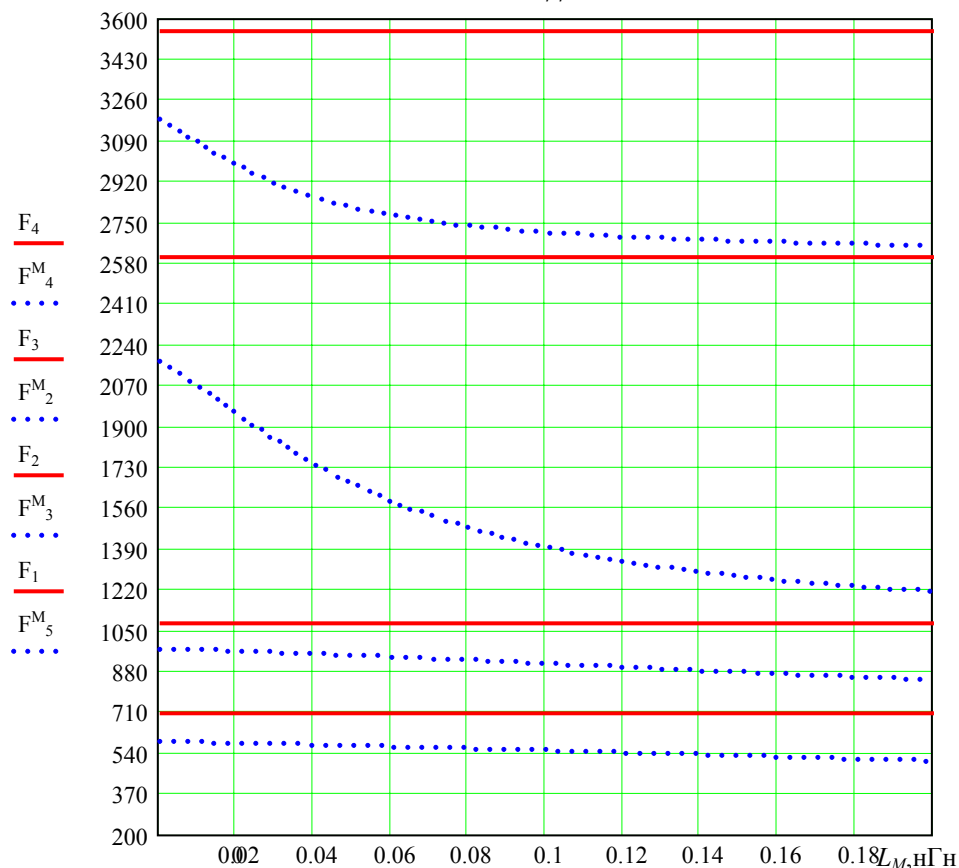


Рисунок 3.2 Изменение частоты формант в зависимости от индуктивности могоаньевого (гортанного) желудочка для звука «А»

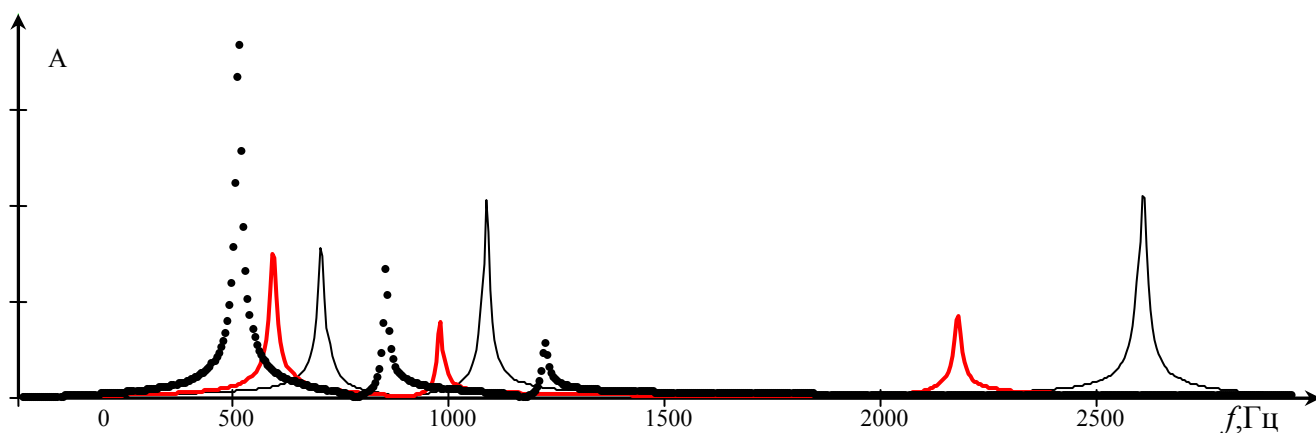


Рисунок 3.3 Амплитуды первых трех формант для звука «А» при разных значениях импеданса голосообразующего тракта

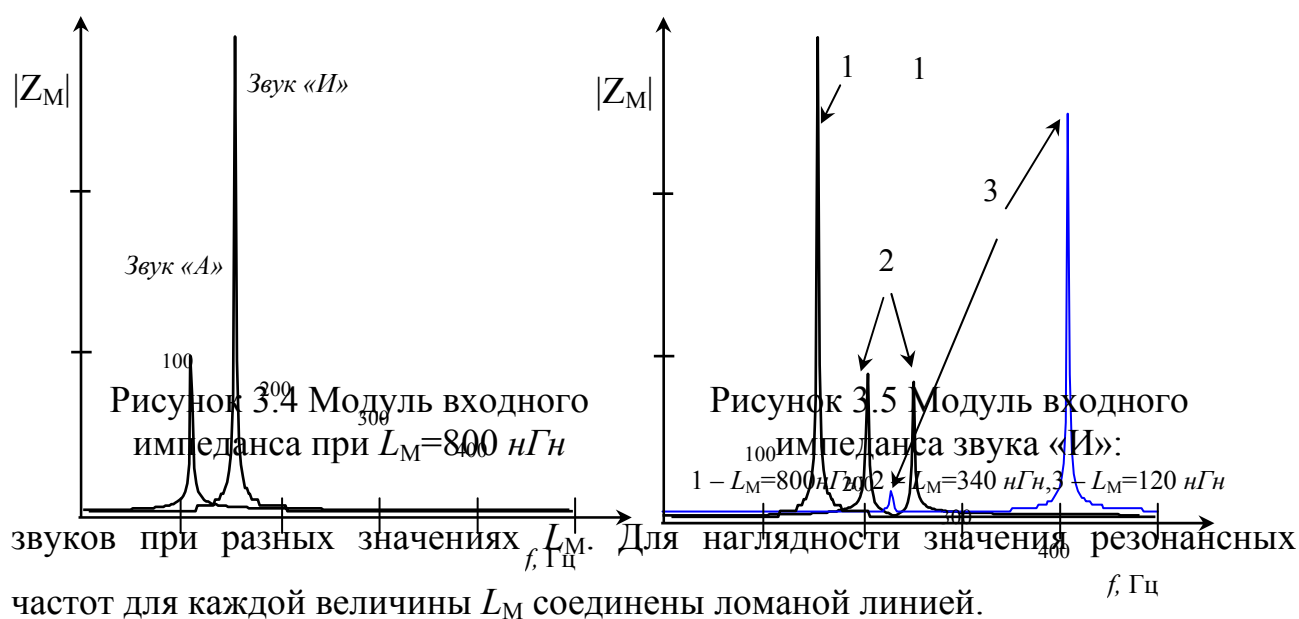
При больших значениях L_M входной импеданс гортанного желудочка Z_M , т.е. нагрузка, на которую работают голосовые складки, может иметь резонанс на частоте, близкой или равной частоте основного тона. Оценки, проведенные на основе этих данных, показывают, что для различных значений L_M указанные резонансы в Z_M в самом деле могут лежать в диапазоне частот основного тона. При этом, если частота первой форманты достаточно низка (звуки «И», «Ы»),

«У»), возможно появление двух резонансов. Результаты расчетов сведены в табл. 3.3.

Таблица 3.3.

$L, \text{ нГн}$	100	200	400	600	800	1000
$F_P^A, \text{ Гц}$	301	221	159	130	113	101
$F_P^O, \text{ Гц}$	281	213	155	129	111	101
$F_P^Y, \text{ Гц}$	364	323	147	121	105	95
	245	197				
$F_P^И, \text{ Гц}$	451	323	255	175	153	137
	233	231	207			
$F_P^Э, \text{ Гц}$	305	223	159	130	113	101
$F_P^Ы, \text{ Гц}$	343	189	135	111	97	87
	241					

На рис. 3.4 приведены модули Z_M для звуков «А» и «И» при $L_M=80 \text{ нГн}$, а на рис. 3.5 для звука «И» при значениях L_M , равных 80 нГн , 34 нГн и 12 нГн . На рис. 3.6 приведены резонансные частоты входного импеданса Z_M для разных



Проведенные оценки приводят к противоречивым результатам. Известно [30], что речевой сигнал практически полностью сформирован на выходе голосовой щели. Следовательно, полюса Z_M должны быть близки к положению формант, а это имеет место или при очень маленьких значениях L_M , или же при значениях $L_M > 100 \text{ нГн}$.

Таким образом, если принять, что индуктивность гортанного желудочка велика, то частота основного тона для одного диктора должна зависеть от звука при его изолированном произнесении. На рис 3.7 представлены экспериментальные значения частоты основного тона реальных речевых сигналов разных дикторов. Данные эксперименты подтверждают гипотезу о возможном влиянии входного импеданса речеобразующего тракта на колебания голосовых складок, а также позволяют указать диапазон возможных значений индуктивности L_M гортанного желудочка. Проведенный анализ показал, что гортанный желудочек может играть существенную роль при формировании звучной речи. Это подтверждается и результатами, приведенными в работе [30], где указывается, что речевой сигнал сформирован на выходе голосовой щели. Косвенно это также подтверждается сравнением рис. 3.6 и 3.7.

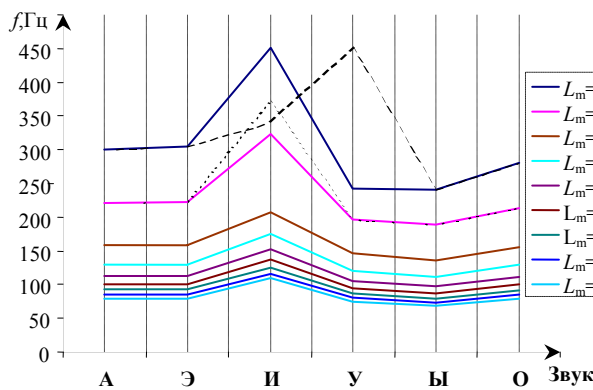


Рисунок 3.6

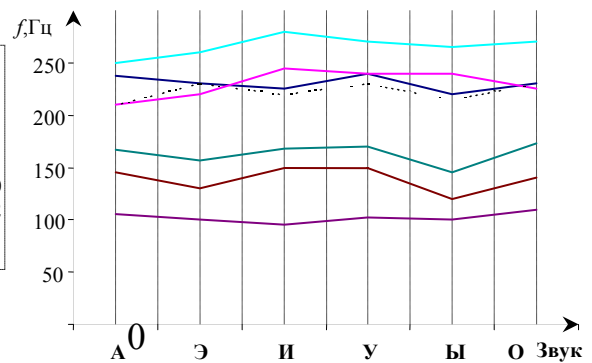


Рисунок 3.7

3.3 Начальные условия

Решение систем уравнений, описывающих образование звучной речи, требует определенных начальных условий, т.е. состояний модели на момент времени $t=0$. Будем считать, что в данный момент времени начинается фонация, а на интервале времени $-\infty < t < 0$ ведется подготовка к этому процессу. Подготовка к началу фонации включает два этапа:

при полном раскрытии голосовых складок увеличение ёмкости C_L и заряд ее до некоторого напряжения $U_L(0)$;

уменьшение ёмкости C_L при полном смыкании голосовых складок, т.е. формировании напряжения $U_L > U_L(0)$, являющегося эквивалентом избыточного давления.

Будем относить момент времени, в котором происходит смыкание голосовых складок, к началу фонации. А также будем считать, что в этот момент времени происходит скачкообразное изменение величины емкости, т.е.

$$C_L = C_L(0) - \Delta C_L \cdot 1(t), \quad (3.15)$$

где ΔC_L – величина, характеризующая уменьшение емкости.

Очевидно, что при данных предположениях соотношение (3.15) будет справедливо и для моделей образования звучной пищеводной речи.

Тогда начальные состояния моделей, представленных на рис. 2.12 и 2.13, т.е. значения токов и напряжений в ветвях и узлах электрических аналогов голосообразующей системы, будут полностью определяться окончанием первого этапа. Основанием к этому служит то, что промежуток времени, в течение которого реализуется этот этап, много больше времени переходных процессов в речеобразующей системе во время вдоха.

Так как площадь раскрытия голосовых складок и псевдоголосовой щели относительно велика, а их продольная длина весьма мала, то на первом этапе можно считать, что

$$R_C = R_\phi \approx 0, \quad L_C = L_\phi \approx 0. \quad (3.16)$$

В начале второго этапа, т.е. при полном их смыкании

$$R_C = R_\phi = \infty, \quad L_C = L_\phi = \infty, \quad (3.17)$$

т.е. происходит полный разрыв электрической цепи. Таким образом, на этапах, предшествующих началу фонации, голосовые складки и псевдоголосовую щель приближенно можно заменить идеальным ключом.

Данные предположения позволяют сделать вывод о том, что начальные условия в основном будут определяться величиной потерь в речеобразующей системе, т.е. значениями R_i и G_i . Ясно, что если R_i и G_i будут определяться соотношениями (2.10), то начальные условия, т.е. распределение токов и напряжений в моделях на рис. 2.12 и 2.13, будут зависеть от генерируемого

звука, так как потери в модели речеобразующей системы на нулевой частоте не равны нулю. Поэтому, чтобы обеспечить независимость начальных условий для моделей (рис. 2.12, 2.13), будем полагать, что

$$R_i = R_j = R_r, \quad G_i = G_j = G_r, \quad (3.18)$$

где R_r и G_r – значения сопротивлений и проводимостей, полученных путем усреднения по всем звукам и всем звеньям.

Тогда, используя (2.5) и (2.10), можно записать, что

$$R_{3B} = b_1 \cdot A \cdot L_{3B}, \quad G_{3B} = b_2 \cdot \frac{C_{3B}}{A}, \quad (3.19)$$

где L_{3B} и C_{3B} – значения индуктивностей и емкостей, полученных путем усреднения данных табл. 2.6 по всем звукам и всем звеньям.

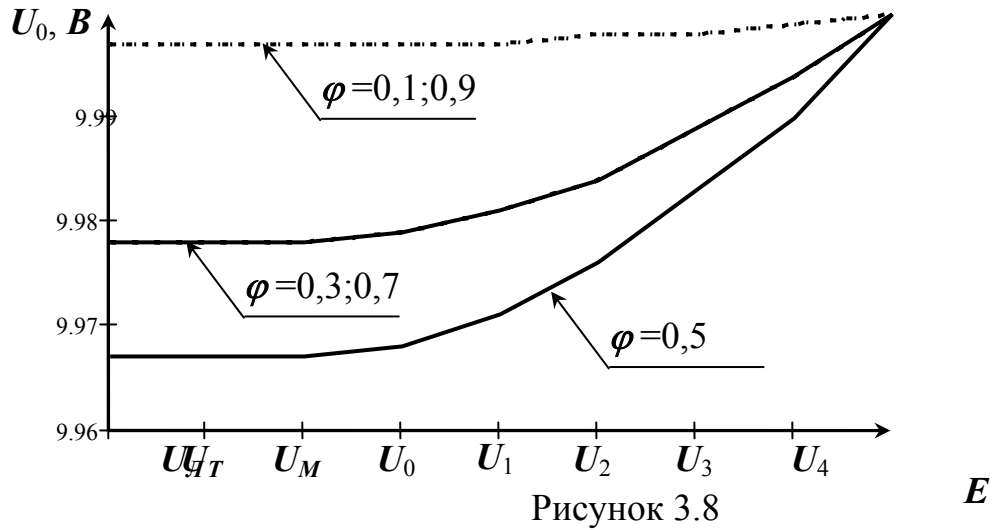
Выражения (3.19) устанавливают связь между параметрами R_{3B} , G_{3B} и A , b_1 и b_2 , запишем их в виде

$$b_1 = \frac{\cos(0,5\pi\varphi)^2}{\tau_{3B}}, \quad b_2 = \frac{\sin(0,5\pi\varphi)^2}{\tau_{3B}}, \quad (3.20)$$

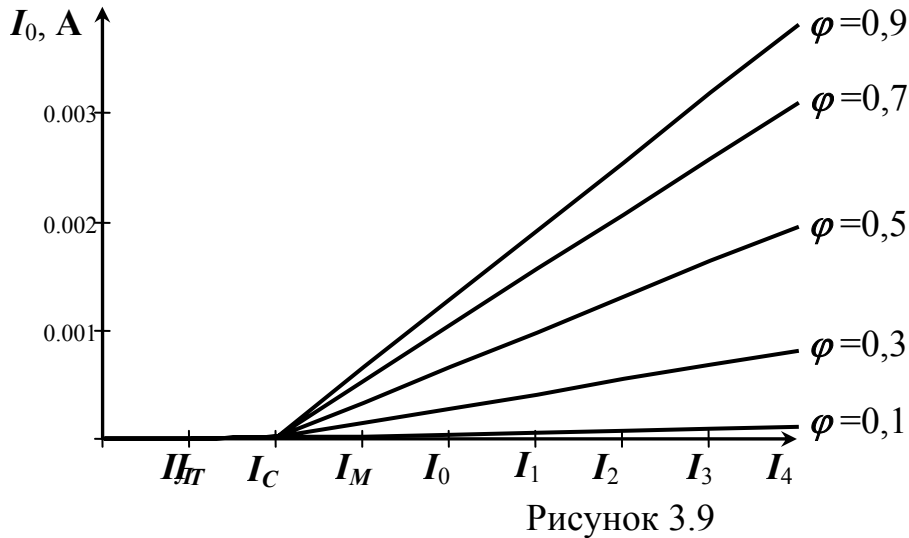
где τ_{3B} – постоянная времени звена модели речеобразующего тракта;
 $0 \leq \varphi \leq 1$.

Аналогичные соотношения будем использовать для оценки потерь в трахее и пищеводе. При изменении φ в заданных пределах b_1 будет изменяться от максимального значения до нуля, а b_2 – наоборот. При этом постоянная времени звена речеобразующего тракта не будет изменяться. Это позволяет исследовать поведение модели системы речеобразования для заданного значения τ_{3B} при различных соотношениях R_{3B} и G_{3B} .

Приведенные соотношения (3.19) и (3.20) позволяют оценить распределение токов и напряжений в моделях речеобразующих систем для образования звуковой нормальной и пищеводной речи. На рис. 3.8 приведены для момента времени $t=0$ значения токов и напряжений в узлах, а на рис. 3.9 – значения токов в ветвях для модели, представленной на рис. 2.12. Данные распределения получены при различных значениях φ для $A=10^7 \text{ Ф}^{-1}$, $E=10 \text{ В}$. Параметры модели были следующие:



постоянная времени звена $\tau_{3B}=0,01$ с;
 постоянная времени трахеи $\tau_T=0,01$ с;
 резонансная частота трахеи $F_T=150$ Гц;
 емкость трахеи $C_T=1$ Ф.



При этом индуктивность трахеи считалась равной:

$$L_T = \frac{1}{C_T (2 \cdot \pi \cdot F_T)^2}. \quad (3.21)$$

Аналогичные результаты получаются и для модели образования пищеводной речи.

Анализ полученных зависимостей показывает, что наибольшее изменение напряжений на узлах модели имеет место при $\varphi=0,5$. Поэтому, учитывая, что фонация определяется избыточным напряжением на емкости C_L , в основном необходимо проанализировать поведение модели при $\varphi=0,5$.

Таким образом, установлены основные соотношения, позволяющие определять начальные условия для систем дифференциальных уравнений, описывающих модели образования звучной нормальной и пищеводной речи. Определены соотношения между сопротивлениями проводимостями звеньев моделей.

3.4 Стационарные решения

В общем виде модели образования нормальной и пищеводной звучной речи можно записать в виде:

$$\frac{dY_H}{dt} = A_H \cdot Y_H, \quad \frac{dY_\Pi}{dt} = A_\Pi \cdot Y_\Pi, \quad (3.22)$$

где A_H и A_Π – матрицы, описывающие образование нормальной и пищеводной звучной речи;

Y_H и Y_Π – векторы параметров.

Стационарные критические точки определяются из решения системы уравнений [53, 104, 105]:

$$A_H \cdot Y_H = 0, \quad A_\Pi \cdot Y_\Pi = 0. \quad (3.23)$$

При анализе уравнений (3.22) и (3.23) будем считать, что избыточное давление равно

$$y_{H0} = U_\Pi = \text{const}, \quad y_{H0} = U_\Pi = \text{const}, \quad (3.24)$$

тогда

$$\frac{U_\Pi}{C_\Pi} \frac{dC_\Pi}{dt} = \frac{i_\Pi}{C_\Pi}, \quad \frac{U_\Pi}{C_\Pi} \frac{dC_\Pi}{dt} = \frac{i_\Pi}{C_\Pi}. \quad (3.25)$$

Будем считать, что емкости C_Π и C_Π изменяются по закону, определяемому соотношением (3.15). Тогда при выполнении условий (2.19) и (2.25) емкости легких и воздушного пузыря будут источниками напряжения с ЭДС, равной

$$U_\Pi = g \cdot U_\Pi(0), \quad U_\Pi = g \cdot U_\Pi(0), \quad (3.26)$$

где $g > 1$ – характеризует скачок напряжения на емкостях $C_{Л}$ и $C_{П}$ перед началом фонации.

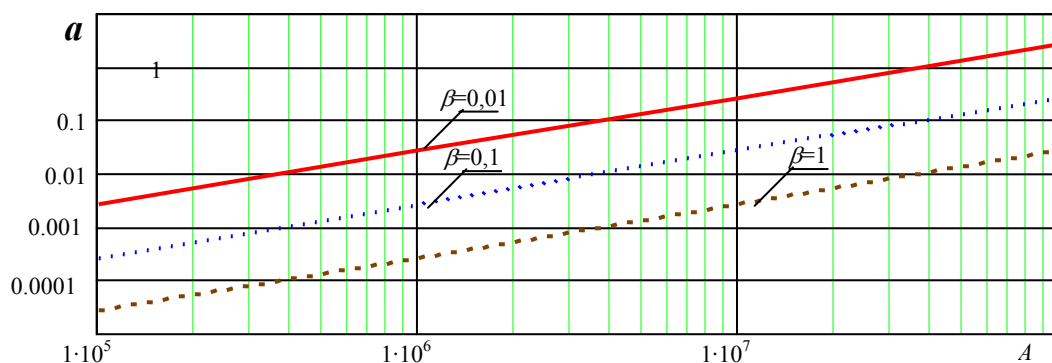


Рисунок 3.10 Зависимость параметра a от A .

Таблица 3.5 Рекуррентные формулы для токов и напряжений в модели речеобразующем тракте

Зависимость	Значения коэффициентов	
$y_{17} = \frac{1}{R_4} y_{16} - \frac{1}{R_4} E$	-	-
$y_{16} = A_0 \cdot y_{15} + B_0 \cdot E$	$B_1 = \frac{B_0}{A_0 + R_3}$	$B_0 = \frac{1}{1 + G_4 \cdot R_4}$
$y_{15} = A_1 \cdot y_{14} + B_1 \cdot E$	$A_1 = \frac{R_4}{A_0 + R_3}$	$B_1 = \frac{B_0}{A_0 + R_3}$
$y_{14} = A_2 y_{13} + B_2 E$	$A_2 = \frac{1}{A_1 + G_3}$	$B_2 = \frac{B_1}{A_1 + G_3}$
$y_{13} = A_3 y_{12} - B_2 \cdot E$	$A_3 = \frac{1}{A_2 + R_2}$	$B_3 = \frac{B_2}{A_2 + R_2}$
$y_{12} = A_4 y_{11} + B_4 \cdot E$	$A_4 = \frac{1}{A_3 + G_2}$	$B_4 = \frac{B_3}{A_3 + G_2}$
$y_{11} = A_5 y_{10} - B_5 \cdot E$	$A_5 = \frac{1}{A_4 + R_1}$	$B_5 = \frac{B_4}{A_4 + R_1}$
$y_{10} = A_6 y_9 + B_6 \cdot E$	$A_6 = \frac{1}{A_5 + G_1}$	$B_6 = \frac{B_5}{A_5 + G_1}$
$y_9 = A_7 y_8 - B_7 \cdot E$	$A_7 = \frac{1}{A_6 + R_0}$	$B_7 = \frac{B_6}{A_6 + R_0}$
$y_8 = A_8 y_7 + B_8 \cdot E$	$A_8 = \frac{1}{A_7 + G_0}$	$B_8 = \frac{B_7}{A_7 + G_0}$
$y_7 = A_9 y_6 - B_9 \cdot E$	$A_9 = \frac{1}{A_8 + R_M}$	$B_9 = \frac{B_8}{A_8 + R_M}$
$y_6 = A_{10} y_5 + B_{10} \cdot E$	$A_{10} = \frac{1}{A_9 + G_M}$	$B_{10} = \frac{B_9}{A_9 + G_M}$

В этом случае модели, описывающие процессы образования нормальной и пищеводной звучной речи, можно рассматривать как автономные системы [70, 71, 89].

Для моделей образования нормальной и пищеводной речи после преобразований (3.23) с учетом соотношений, приведенных в табл. 2.7 и 2.8 получаем уравнения для тока через голосовые складки:

$$\begin{aligned} \gamma \cdot \lambda \cdot y_{H5}^4 + (\gamma + \lambda) \cdot R \cdot y_{H5}^3 + (R^2 - \gamma \cdot Z - \lambda \cdot Z) \cdot y_{H5}^2 - (2 \cdot R \cdot Z + a \cdot \beta \cdot \omega_C^2) \cdot y_{H5} + Z^2 = 0; \quad (3.27) \\ \gamma \cdot \lambda \cdot y_{П5}^4 + (V \cdot \lambda + \gamma \cdot R_g + \gamma \cdot V) \cdot y_{П5}^3 + (V^2 - \gamma \cdot W + V \cdot R_g - W \cdot \lambda) \cdot y_{П5}^2 - \\ - (2W \cdot V + W \cdot R_g + a \cdot \beta \cdot \omega_C^2) \cdot y_{П5} + W^2 = 0 \end{aligned}$$

$$\text{где } Z = \frac{g_T U_L(0)}{1 + G_T R_T} - B_{10} \cdot E; \quad R = \frac{R_T}{1 + G_T R_T} - A_{10};$$

$$W = \frac{g_{П} U_{П}(0)}{1 + G_{П} R_{П}} - B_8 \cdot E; \quad V = \frac{R_{П}}{1 + G_{П} R_{П}} - A_8;$$

y_{H5} — ток через голосовые складки для модели образования нормальной звучной речи;

$y_{П5}$ — ток через голосовые складки для модели образования пищеводной речи.

Значения для коэффициентов A_{10} и B_{10} приведены в табл. 3.5. В ней приведены также рекуррентные формулы для расчета токов и напряжений в электрическом аналоге речеобразующего тракта. При этом необходимо учитывать, что в модели образования пищеводной речи гортанный желудочек, т.е. C_M и L_M , удален.

Численное решение уравнений (3.27) дает четыре различных корня, два из которых комплексные, один отрицательный и один положительный. Так как скорость воздушного потока через голосовые складки характеризуется положительным действительным числом [108, 116, 125, 127], то достаточно провести исследование только этого решения.

Отличие процесса образования пищеводной речи состоит только в другой системе образования избыточного давления и месте формирования аналога голосовой щели [37, 38, 59, 90, 97, 129]. Однако необходимо учитывать, что в процессе становления пищеводной речи [121] эти части подстраиваются под существующий речеобразующий тракт [8, 9, 13, 14, 97]. Следовательно, можно

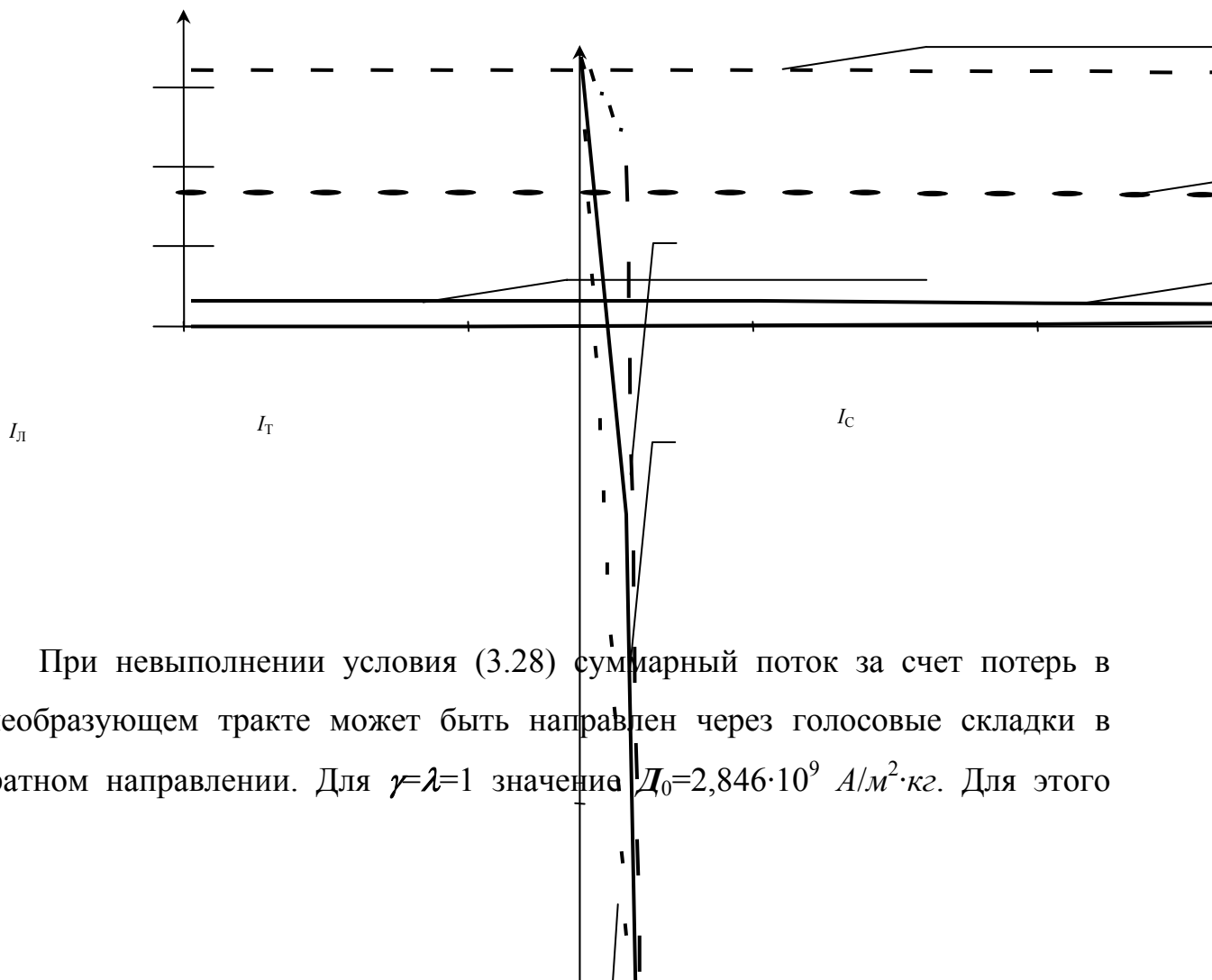
ожидать, что система генерации пищеводной звучной речи стремится приблизиться по своим параметрам к системе генерации нормальной звучной речи. Поэтому основное внимание при исследовании стационарных решений было уделено модели образования нормальной речи.

Из табл. 3.5 следует, что параметры A_i и B_i зависят от масштабного коэффициента A . Следовательно, решения уравнений (3.27) зависят от значения параметров A , λ , γ , β , a , ω_C , τ_{3B} и τ_T или τ_H . При малых значениях λ и γ порядка 1, влияние сил Бернулли на движение голосовых складок и ток через голосовые складки практически равны нулю. Можно считать, что при значениях $\gamma=\lambda=1$ решения являются граничными, т.е. при которых силы Бернулли не влияют на процессы в голосовой щели.

Численное моделирование показало, что параметры β , a и A связаны соотношением

$$D = \frac{A}{a \cdot \beta} > D_0, \quad (3.28)$$

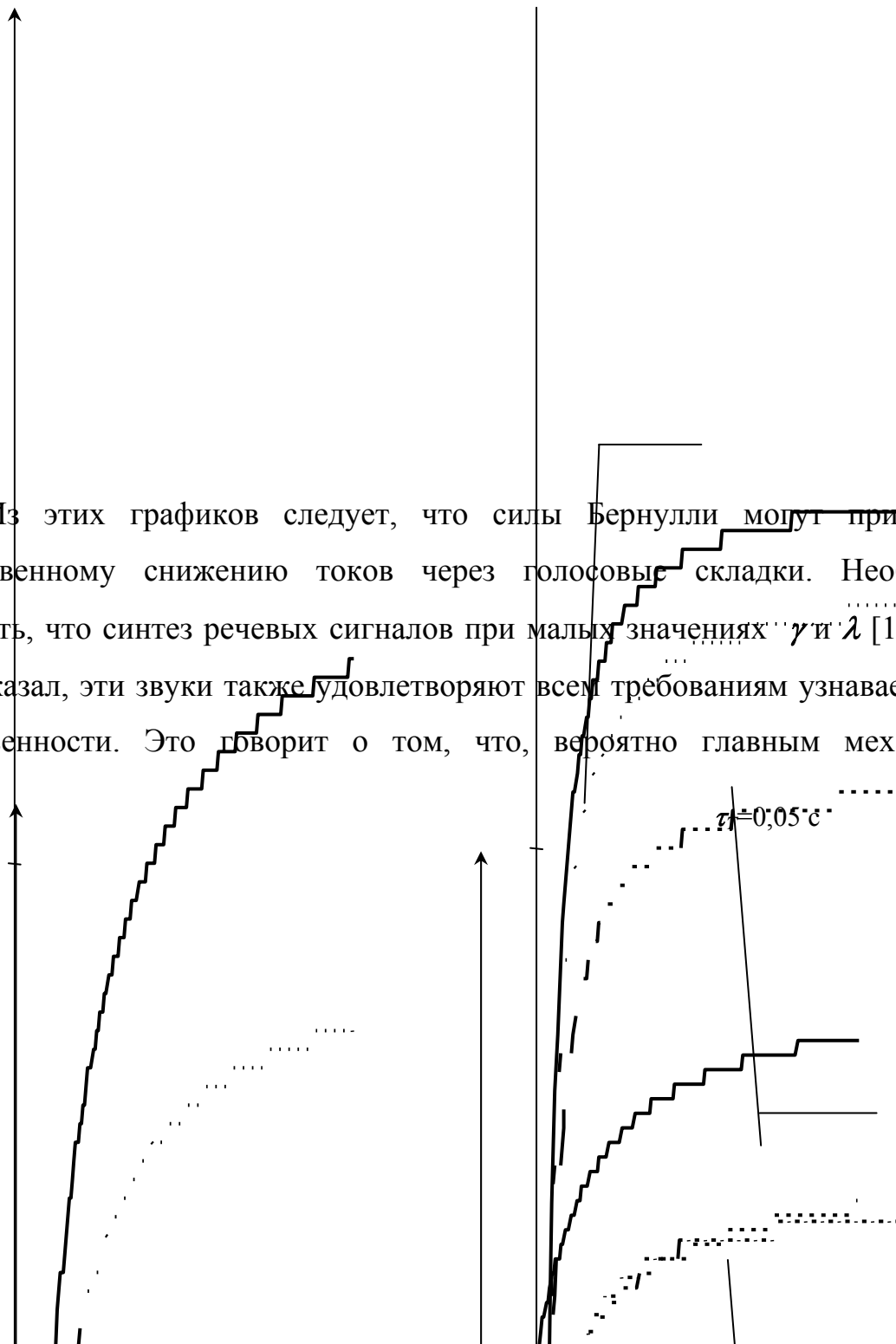
где D_0 – некоторая граничная величина, слабо зависящая от γ и λ .



При невыполнении условия (3.28) суммарный поток за счет потерь в речеобразующем тракте может быть направлен через голосовые складки в обратном направлении. Для $\gamma=\lambda=1$ значения $D_0=2,846 \cdot 10^9 \text{ А/м}^2 \cdot \text{кг}$. Для этого

значения D_0 ток на выходе речеобразующего тракта равен нулю, т. е. ток за счет избыточного давления и ток потерь взаимно уравниваются друг друга. На рис. 3.10 приведены зависимости параметра a от A для трех значений β при $\lambda=\gamma=1$. На рис. 3.11 и 3.12 приведено распределение токов и напряжений в модели системы речеобразования для стационарного режима при $E=10$; $g=1,2$; $\varphi=0,5$; $\beta=1,25 \cdot 10^{-3}$; $\omega_c=2\pi 150$; $\tau_{зв}=\tau_l=0,025$ с. На этих рисунках кривые 1 – характеризуют начальные условия, 2 – стационарные решения при $D=D_0$ и $\gamma=\lambda=1$, 3 – при $D=10^{11}$ и $\gamma=\lambda=1$ и 4 – при $D=10^{11}$ и $\gamma=\lambda=10^4$.

Из этих графиков следует, что силы Бернулли могут привести к существенному снижению токов через голосовые складки. Необходимо отметить, что синтез речевых сигналов при малых значениях γ и λ [14,15, 58, 59] показал, эти звуки также удовлетворяют всем требованиям узнаваемости и естественности. Это говорит о том, что, вероятно главным механизмом



управления голосовых складок является перепад давлений.

Основным параметром, определяющим процесс фонации, является площадь раскрытия голосовых складок. На рис. 3.13 и 3.14 приведены графики площади раскрытия голосовых складок в зависимости от постоянной времени звена речеобразующего тракта. При расчетах использовались следующие значения параметров: $A=10^7$; $D=10^{11}$; $\omega_c=2\pi 150$ Гц. Полученные значения максимальной площади раскрытия голосовых складок соответствуют ее реальному значению (3.2), что позволяет принять значения данных параметров за основу. На рис. 3.15 и 3.16 приведены графики токов через голосовые складки, рассчитанные для тех же значений параметров. Сравнение этих рисунков показывает, что силы Бернулли несущественно влияют на степень раскрытия голосовых складок, но значительно ограничивают ток через голосовые складки.

На рис. 3.18 приведена зависимость площади раскрытия голосовых складок от параметра β , а на рис. 3.19 – площади, тока и сопротивления голосовой щели в зависимости от параметра a . С учетом соотношения (3.2) эти графики позволяют дать оценки основных параметров модели речеобразующей системы при $\gamma=\lambda=1$.

Ясно, что с ростом значений γ и λ может наступить такой момент, когда произойдет полный срыв колебаний голосовых складок. На рис. 3.20 приведены граничные значения $\gamma=\lambda$, а на рис. 3.21 - площадь раскрытия голосовой щели для этих значений γ и λ в зависимости от D для двух значений A при $\beta=0,01$. На рис. 3.22 для этих же условий приведена зависимость отношения падения напряжения за счет сил Бернулли к падению напряжения на сопротивлении голосовой щели. Эта зависимость хорошо коррелируется со значением коэффициента Бернулли [117], что косвенно подтверждает адекватность модели.

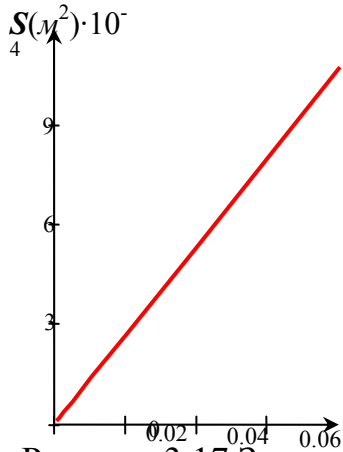


Рисунок 3.17 Зависимость площади раскрытия голосовых складок от β раскрытия голосовых складок при $D=10^{11}$; $\gamma=\lambda=1$; $A=10^7$

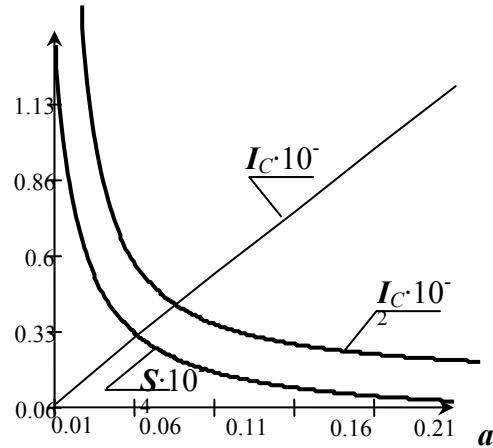


Рисунок 3.18 Зависимости площади раскрытия голосовых складок, тока через голосовой щели от параметра a

Графики на рис. 3.10 – 3.21 получены при резонансной частоте голосовых

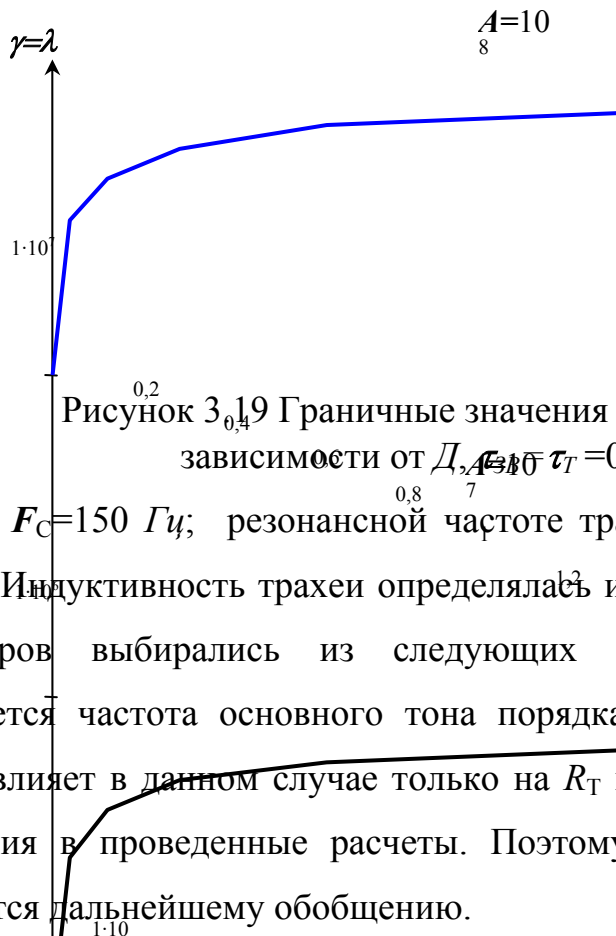


Рисунок 3.19 Граничные значения параметра $\gamma=\lambda$ в зависимости от D , A и $\tau_T=0,25 \cdot 10^{-3} \text{ с}$

складок $F_C=150 \text{ Гц}$; резонансной частоте трахеи $F_T=150 \text{ Гц}$; емкости трахеи $C_T=1 \text{ Ф}$. Индуктивность трахеи определялась из условия (3.21). Такие значения параметров выбирались из следующих соображений. Наиболее часто встречается частота основного тона порядка 150 Гц . Резонансная частота трахеи влияет в данном случае только на R_T и G_T , которые вносят небольшие изменения в проведенные расчеты. Поэтому полученные результаты легко поддаются дальнейшему обобщению.

Таким образом, в результате анализа стационарных решений установлена связь (3.28) между параметром A , определяющим соотношение между значением емкостей и индуктивностей речеобразующего тракта, параметром a ,

определяющем влияние разности напряжений на голосовые складки, и параметром β , определяющим связь между площадью раскрытия голосовой щели и ее сопротивлением.

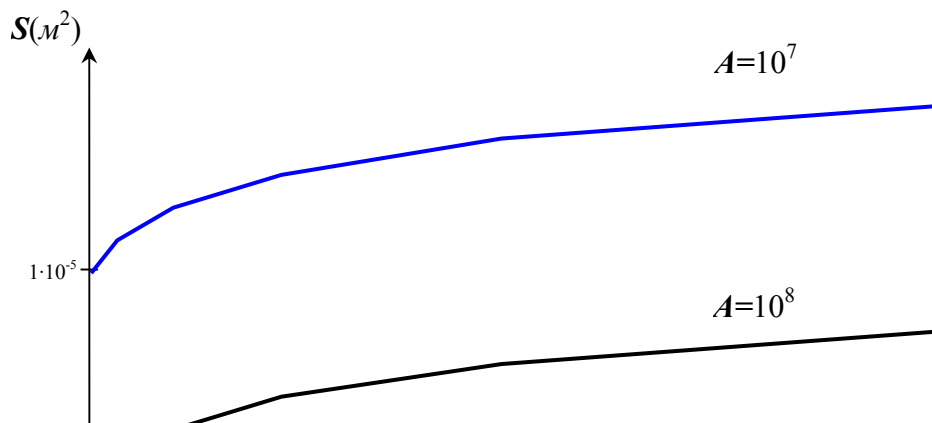


Рисунок 3.20 Площадь раскрытия голосовых складок для граничных значений $\gamma=\lambda$; $\tau_{3B}=\tau_T=0,025c$

Кроме того, установлена верхняя граница значений γ и λ , определяющая влияние сил Бернулли на движение голосовых складок и ток через них.

Полученные результаты коррелируются с данными по максимальной площади раскрытия голосовой щели, что косвенно подтверждает адекватность предложенной модели речеобразующей системы.

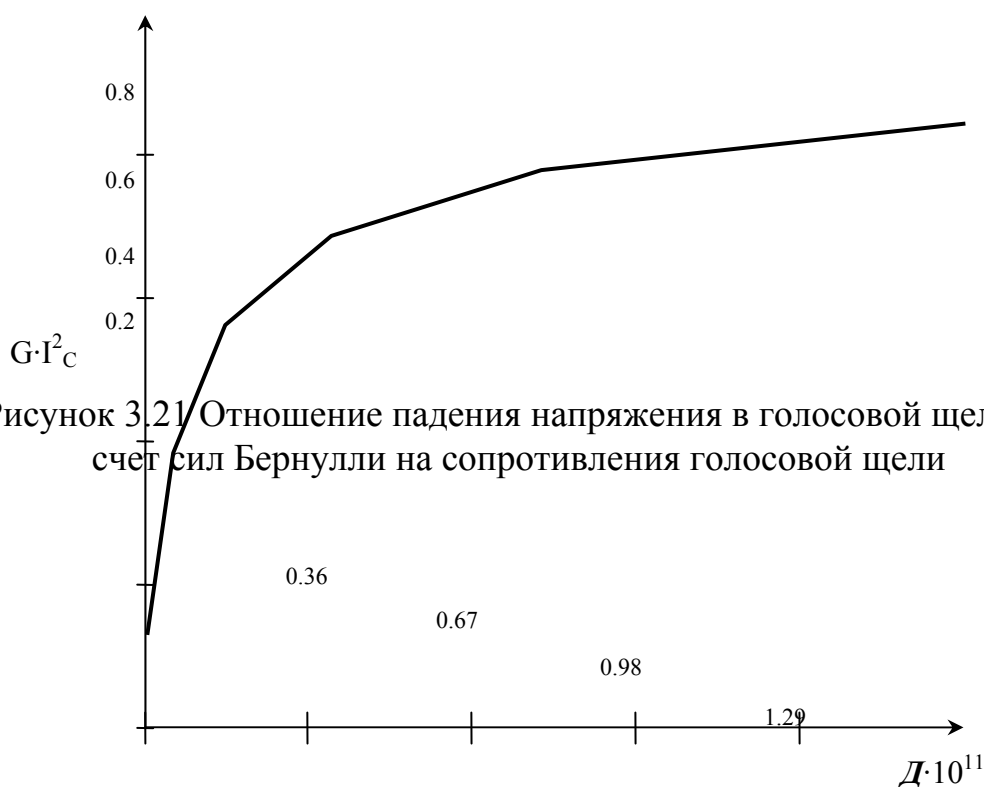


Рисунок 3.21 Отношение падения напряжения в голосовой щели за счет сил Бернулли на сопротивления голосовой щели

3.5 Исследование устойчивости процессов образования звучной речи

Исследования устойчивости проводились в стационарных критических точках, в которых решения уравнений (3.27) определялись действительными положительными значениями площади раскрытия голосовой щели и тока через нее. Линеаризованные матрицы A_H и A_{II} для этого случая представлены (3.30) и (3.31).

$$A_H = \begin{pmatrix} -\frac{R_H}{L_H} & -\frac{1}{L_H} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{C_H} & -\frac{G_H}{C_H} & 0 & 0 & -\frac{1}{C_H} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{a} & -\omega_c^2 & -\delta & -2\frac{rY_5}{a} & \frac{1}{a} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_2 & x_3 & x_4 & x_5 & x_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{C_0} & -\frac{G_0}{C_0} & -\frac{1}{C_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{L_0} & \frac{R_0}{L_0} & -\frac{1}{L_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_1} & -\frac{G_1}{C_1} & -\frac{1}{C_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_1} & -\frac{R_1}{L_1} & -\frac{1}{L_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_2} & -\frac{G_2}{C_2} & -\frac{1}{C_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_2} & -\frac{R_2}{L_2} & -\frac{1}{L_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_3} & -\frac{G_3}{C_3} & -\frac{1}{C_3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_3} & -\frac{R_3}{L_3} & -\frac{1}{L_3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_4} & -\frac{G_4}{C_4} & -\frac{1}{C_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_4} & -\frac{R_4}{L_4} \end{pmatrix} \quad (3.30)$$

где

$$x_2 = \frac{Y_3}{\alpha - L_g \cdot Y_3};$$

$$x_3 = \left(\frac{Y_2 - Y_6}{\alpha + L_g \cdot Y_3} - \frac{\lambda \cdot Y_3^2}{\alpha + L_g \cdot Y_3} \right) \cdot \left(1 - \frac{L_g \cdot Y_3}{\alpha + L_g \cdot Y_3} \right) - \frac{Y_4 \cdot Y_5}{Y_3(\alpha + L_g \cdot Y_3)} \cdot \left(\frac{1}{Y_3} - \frac{L_g}{\alpha + L_g \cdot Y_3} \right) +$$

$$+ \frac{Y_5}{\alpha + L_g \cdot Y_3} \cdot \left(\frac{\beta \cdot L_g}{\alpha + L_g \cdot Y_3} + \frac{L_g \cdot R_g \cdot Y_3}{\alpha + L_g \cdot Y_3} - R_g \right);$$

$$x_4 = \frac{\alpha \cdot Y_5}{Y_3(\alpha + L_g \cdot Y_3)};$$

$$x_5 = \frac{\alpha \cdot Y_4}{Y_3(\alpha + L_g \cdot Y_3)} - 2 \frac{\lambda \cdot Y_3 \cdot Y_5}{\alpha + L_g \cdot Y_3} - \frac{\beta \cdot R_g \cdot Y_3}{\alpha + L_g \cdot Y_3};$$

$$x_6 = \frac{Y_3}{\alpha + L_g \cdot Y_3}.$$

$$A_{II} = \begin{vmatrix} -\frac{R_T}{L_T} & -\frac{1}{L_T} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{C_T} & -\frac{G_T}{C_T} & 0 & 0 & -\frac{1}{C_T} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{a} & -\omega_C^2 & -\delta & -\frac{2 \cdot Y \cdot Y_5}{a} & -\frac{1}{a} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{Y_3}{\alpha} & x_3 & \frac{Y_5}{Y_3} & x_5 & -\frac{Y_3}{\alpha} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{C_M} & -\frac{G_M}{C_M} & -\frac{1}{C_M} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{L_M} & -\frac{R_M}{L_M} & -\frac{1}{L_M} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_0} & -\frac{G_0}{C_0} & -\frac{1}{C_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_0} & -\frac{R_0}{L_0} & -\frac{1}{L_0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_1} & -\frac{G_1}{C_1} & -\frac{1}{C_1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_1} & -\frac{R_1}{L_1} & -\frac{1}{L_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_2} & -\frac{G_2}{C_2} & -\frac{1}{C_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_2} & -\frac{R_2}{L_2} & -\frac{1}{L_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_3} & -\frac{G_3}{C_3} & -\frac{1}{C_3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_3} & -\frac{R_3}{L_3} & -\frac{1}{L_3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{C_4} & -\frac{G_4}{C_4} & -\frac{1}{C_4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{L_4} & -\frac{R_4}{L_4} \end{vmatrix} \quad (3.31)$$

где

$$x_3 = \frac{Y_2 - Y_6}{\alpha} - \frac{Y_4 - Y_5}{Y_3^2} - \frac{\lambda \cdot Y_5^2}{\alpha} ;$$

$$x_5 = \frac{Y_4}{Y_3} - \frac{\beta}{\alpha} - 2 \cdot \frac{\lambda \cdot Y_4 \cdot Y_5}{\alpha} ;$$

$$x_5 = \frac{\alpha \cdot Y_4}{Y_3(\alpha + L_g \cdot Y_3)} - 2 \frac{\lambda \cdot Y_3 \cdot Y_5}{\alpha + L_g \cdot Y_3} - \frac{\beta \cdot R_g \cdot Y_3}{\alpha + L_g \cdot Y} ;$$

$Y_2; Y_3; Y_4; Y_5; Y_6$ – значения соответствующих переменных в стационарных критических точках.

Анализ устойчивости проводился численными методами в среде MathCAD [111]. Определялись собственные значения соответствующих матриц и анализировались их действительные части. При этом считалось, что если все действительные части собственных значений отрицательны, то система имеет устойчивые колебательные решения [10, 70, 71].

На рис. 3.22 приведены области устойчивых решений для моделей образования звучной нормальной речи в зависимости от постоянной времени трахеи и звеньев речеобразующего тракта при разных значениях постоянной времени голосовой щели τ_C . Расчеты проводились при следующих параметрах: $D=10^{11}$; $\beta=1,25 \cdot 10^{-3}$; $A=10^7$; $a=0,08$; $\varphi=0,5$; $F_T=150 \text{ Гц}$; $F_C=150 \text{ Гц}$; $C_t=1$; $L_M=10^{-9}$.

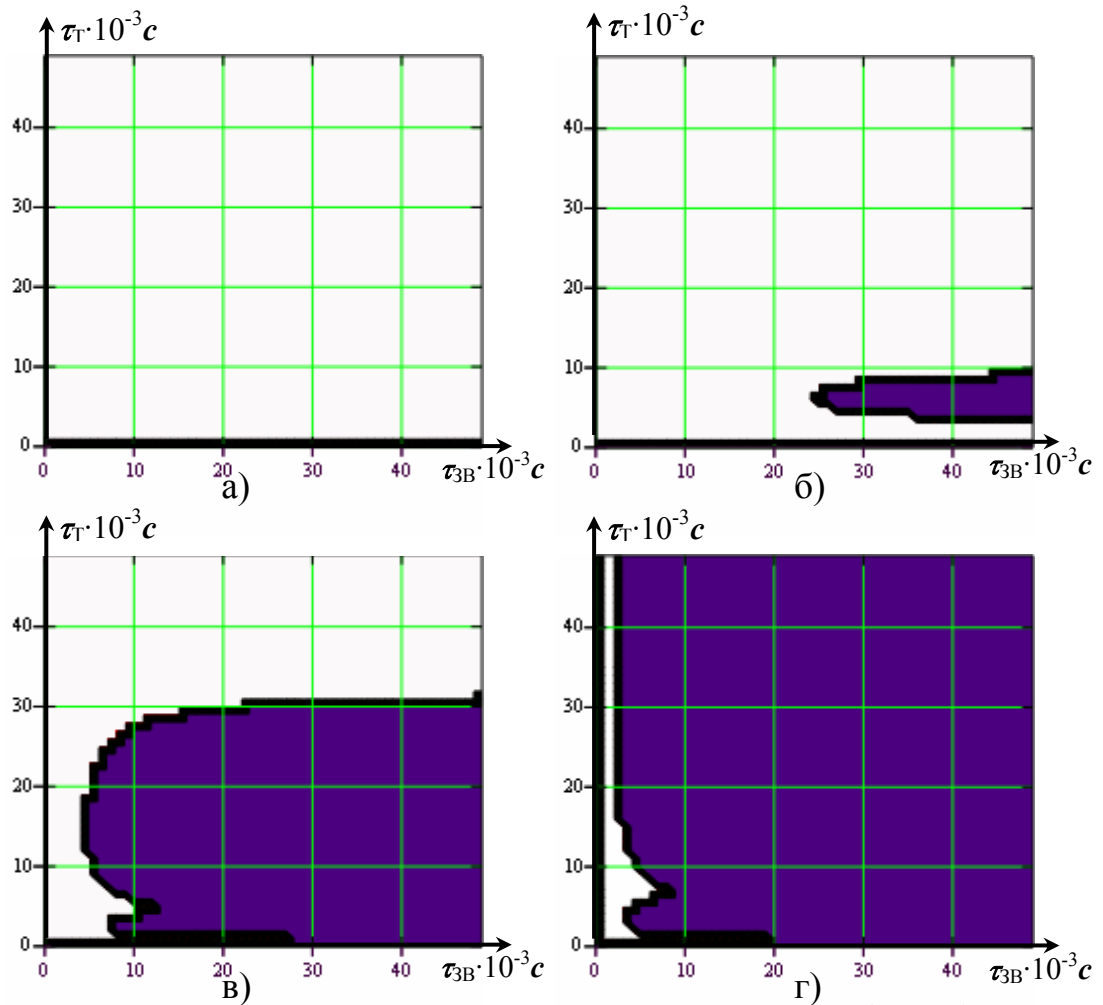


Рисунок 3.22 Зоны устойчивых колебаний (белое) в зависимости от τ_T и τ_{3B} при $C_T=1$ Φ (а - $\tau_C=1 \cdot 10^{-3} c$; б - $\tau_C=1,25 \cdot 10^{-3} c$, в - $\tau_C=1,5 \cdot 10^{-3} c$; г - $\tau_C=1,75 \cdot 10^{-3} c$)

Было обнаружено, что области устойчивых решений не зависят от параметров λ и γ в диапазоне их значений от 1 до 10^4 . Из рис. 3.22 следует, что при $\tau_C < 10^{-3} c$ в заданном диапазоне изменения τ_T и τ_{3B} система полностью устойчива. Это хорошо коррелируется с данными по постоянной времени голосовой щели, приведенными в работах [117, 125, 127].

На рис. 3.23 приведены зоны устойчивых решений зависимости от постоянных времени τ_T и τ_{3B} при $\tau_C=0,25 \cdot 10^{-3} c$ для различных значений емкости трахеи C_T . При этом было обнаружено, что эти зоны не зависят от значений индуктивности L_M . Из этого рисунка следует, что при значениях $\tau_{3B} > 0,01 c$, а $\tau_T > 0,01 c$ колебания в системе становятся устойчивыми практически в широком диапазоне изменений параметров C_T , L_M , λ и γ .

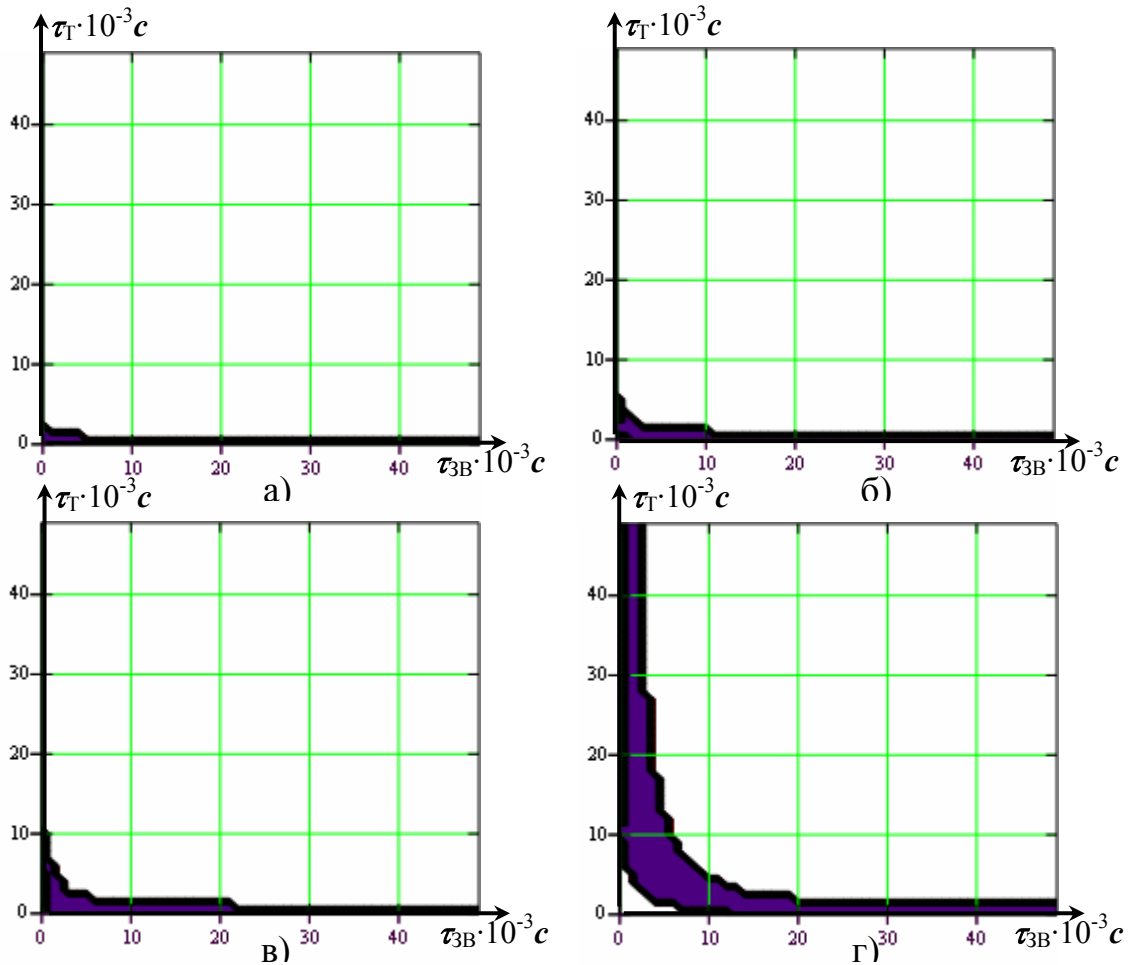


Рисунок 3.23 Зоны устойчивых колебаний (белое) в зависимости от τ_{3B} и τ_T при $\tau_C=0,25 \cdot 10^{-3} c$ (а - $C_T=40 \Phi$; б - $C_T=80 \Phi$, в - $C_T=160 \Phi$; г - $C_T=320 \Phi$)

На рис. 3.24 приведены области устойчивых колебаний в зависимости от емкости C_T и резонансной частоты F_T при различных значениях τ_{3B} и τ_T , из которых следует, что при $\tau_{3B} > 0,01 c$, $\tau_T > 0,01 c$ и $\tau_C = 0,25 \cdot 10^{-3} c$ система также характеризуется устойчивыми колебаниями при изменении параметров F_T и C_T в широких пределах.

Проведенный анализ показывает, что оценки, данные в разделе 3.1, соответствуют устойчивым режимам работы модели образования звучной нормальной речи. При этом во всех расчетах значения δ выбирались в соответствии с (3.7). При этом необходимо учитывать, что значения τ_{3B} следует выбирать не менее $0,02-0,03 c$, а $\tau_C = 0,25 \cdot 10^{-3} c$, это также соответствует соотношениями (3.39) и (3.10).

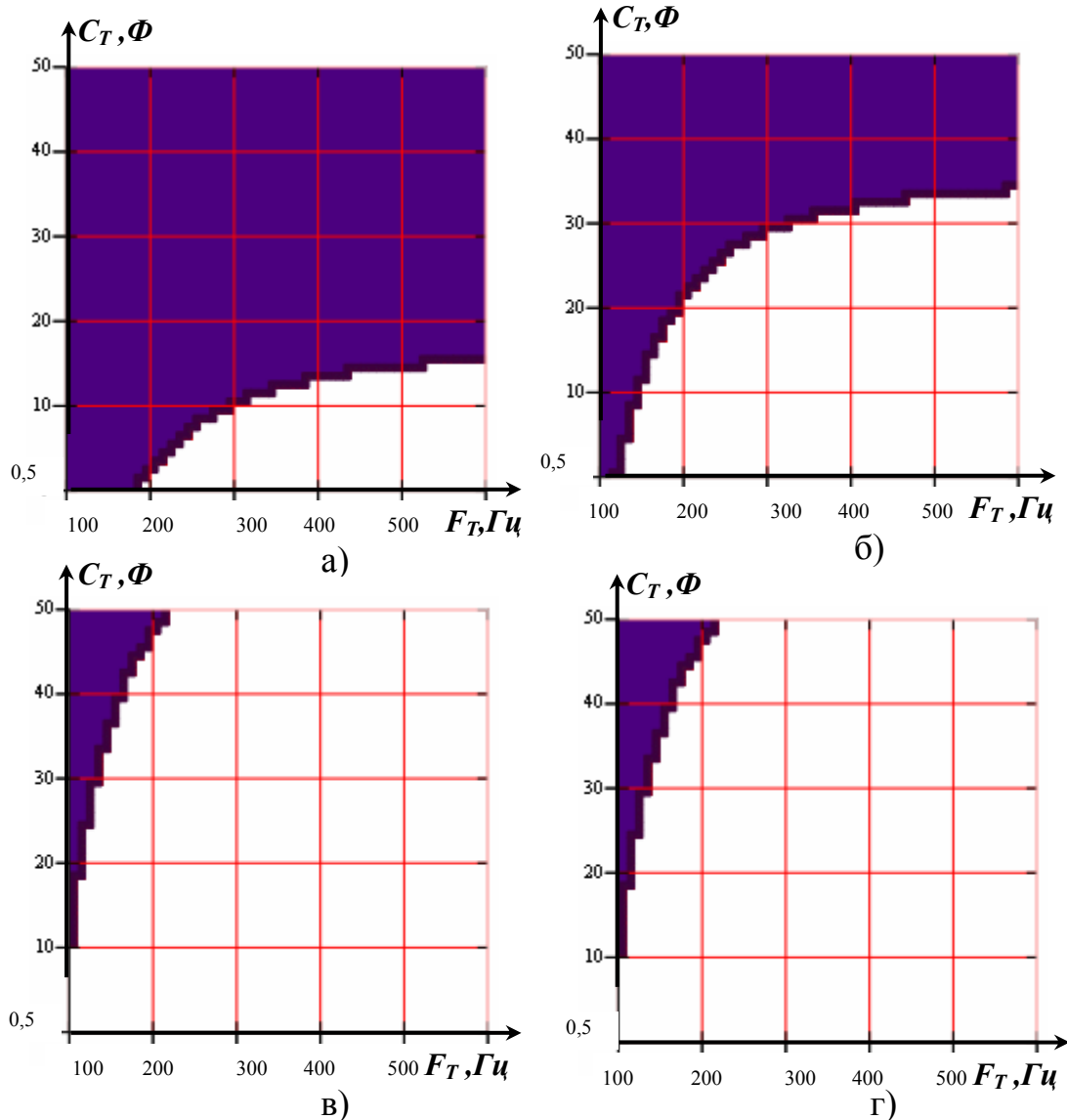


Рисунок 3.24 Зоны устойчивых колебаний (белое) в зависимости от F_T и C_T при $\tau_C=0,25$ мс, (а - $\tau_{ЗВ}=\tau_{Г}=0,001$ с; б - $\tau_{ЗВ}=\tau_{Г}=0,0015$ с; $L_M=10^{-9}$ Гн; в - $\tau_{ЗВ}=\tau_{Г}=0,002$ с; $L_M=10^{-9}$ Гн ; г - $\tau_{ЗВ}=\tau_{Г}=0,002$ с; $L_M=10^{-8}$ Гн)

На рис. 3.25 приведены зависимости собственных, т.е. резонансных, частот модели речеобразующей системы при раскрытых голосовых складках от резонансной частоты трахеи при двух значениях емкости трахеи для звука «А». Аналогичные графики получены и для других звуков.

В табл. 3.8 эти резонансные частоты приведены для всех звуков при значениях F_T равных 100 и 600 Гц, и емкости трахеи – 5,5 и 45,5 Φ .

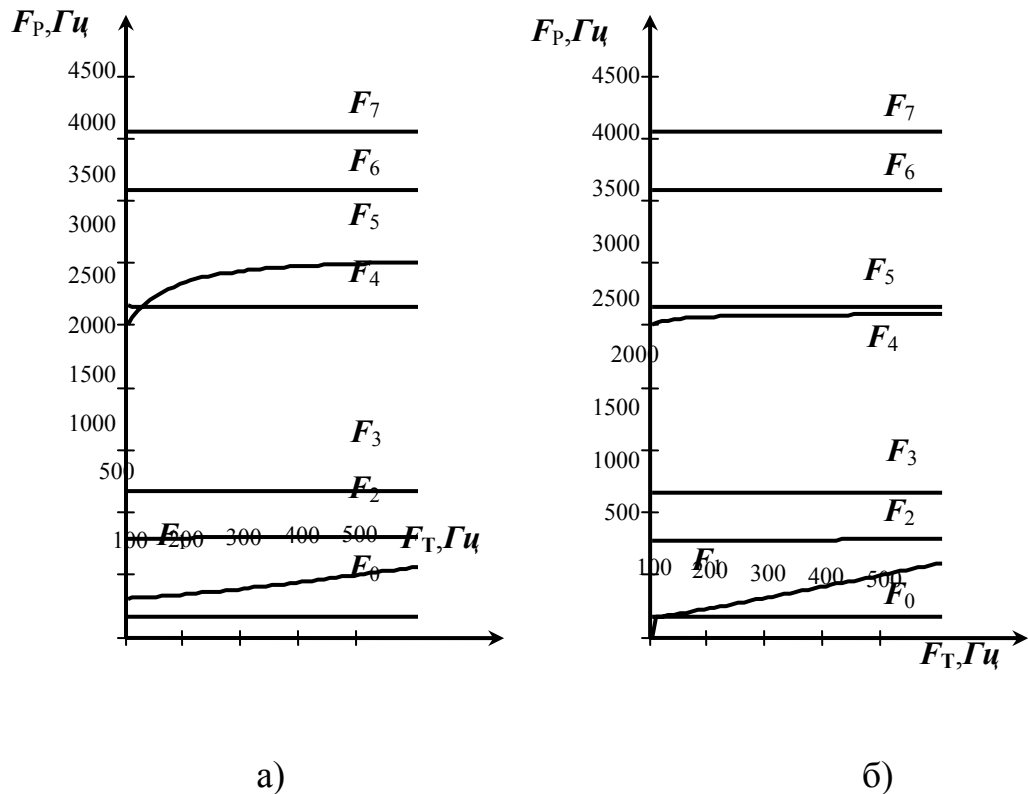


Рисунок 3.25 Зависимость резонансной частоты речеобразующей системы о частоты F_T , а) – при $C_T=5\Phi$; б) – при $C_T=45,5\Phi$

Сравнение данных, приведенных в табл. 2.3, с данными табл. 3.8 показывает, что эти резонансные частоты коррелированы с частотами формант, но появляются дополнительные резонансные частоты. По-видимому, этим обстоятельством объясняются появления так называемых ложных формант [86].

Исследование областей устойчивых решений для моделей образования пищеводной речи проводилось в зависимости от постоянных времени τ_{3B} , τ_P , C_P , F_P , а также постоянных времени

$$\tau_g = \frac{L_g}{R_g} \quad \text{и} \quad \tau_s = \frac{\alpha}{\beta}.$$

При этом остальные параметры выбирались так же, как и для моделей образования нормальной звуочной речи.

На рис. 3.26 представлены области устойчивых решений в зависимости от τ_{3B} и τ_P при различных значениях τ_g , τ_s и F_P . На рис. 3.27 - зоны устойчивых колебаний в зависимости от частоты F_P и емкости C_P при различных значениях τ_P . Из этих рисунков следует, что зоны устойчивых колебаний более

узки, чем для моделей образования звучной нормальной речи. При этом они очень критичны к постоянным времени τ_g и τ_s и значения их должны быть на один - два порядка выше, чем для моделей образования нормальной звучной речи.

Таблица 3.8

Звук	F_T	C_T	F_0	F_1	F_2	F_3	F_4	F_5	F_6	F_7
А	100	5,5	165,0	310,8	763,5	1172	2657	2510	3587	4057
	100	45,5	162,2	Нет	780,0	112	2654	2512	3587	4057
	600	5,5	162,5	564	807,1	1178	2649	3009	3587	4058
	600	45,5	162,5	593	783	1163	2654	2589	3587	4057
О	100	5,5	164,9	285,1	649,5	1103	2525	2679	3580	4008
	100	45,5	165,3	150,7	635,5	1045	2541	2673	3581	4011
	600	5,5	162,8	512,5	670,4	1135	2632	3045	3581	4013
	600	45,5	163,0	575,7	647,8	1053	2657	2627	3581	4011
У	100	5,5	164,9	235,2	479,5	854,2	2453	2580	3450	4027
	100	45,5	166,8	147,1	413,2	777,2	2426	2579	3450	4028
	600	5,5	163,7	348,1	571,1	911,0	2561	2933	3451	4028
	600	45,5	164,7	389,4	590,8	795,8	2574	2509	3450	4028
И	100	5,5	165,1	Нет	619,7	2268	3209	3337	3585	4425
	100	45,5	169,8	Нет	376,9	2267	3208	3582	3556	4426
	600	5,5	163,8	286,4	813,2	2270	3209	3578	3928	4426
	600	45,5	165,4	325,6	636,8	2268	3208	3580	3654	4426
Э	100	5,5	165,3	355,4	646,1	1824	2571	2581	3430	4179
	100	45,5	170,4	Нет	530,7	1821	2570	2601	3430	4179
	600	5,5	162	451,4	779,3	1826	2567	3059	3430	4179
	600	45,5	160,1	494,8	641,8	1822	2574	2674	3430	4179
Ы	100	5,5	169,3	251,6	590,8	1517	2360	2311	3226	4239
	100	45,5	163,4	0	430,6	1512	2361	2240	3227	4239
	600	5,5	164,5	337,1	731,5	1521	2355	2757	3227	4239
	600	45,5	160,1	385,9	629,0	1513	2359	2304	3227	4239

Таким образом, анализ областей возникновения устойчивых колебаний в моделях образования нормальной и пищеводной речи позволяет уточнить значения параметров соответствующих электрических аналогов. При этом полученные значения этих параметров коррелируются с данными, полученными другими методами [116, 117, 125, 127, 146, 151, 155, 156, 159, 162, 169, 170, 171], это позволяет сделать косвенный вывод об адекватности модели соответствующим реальным системам речеобразования.

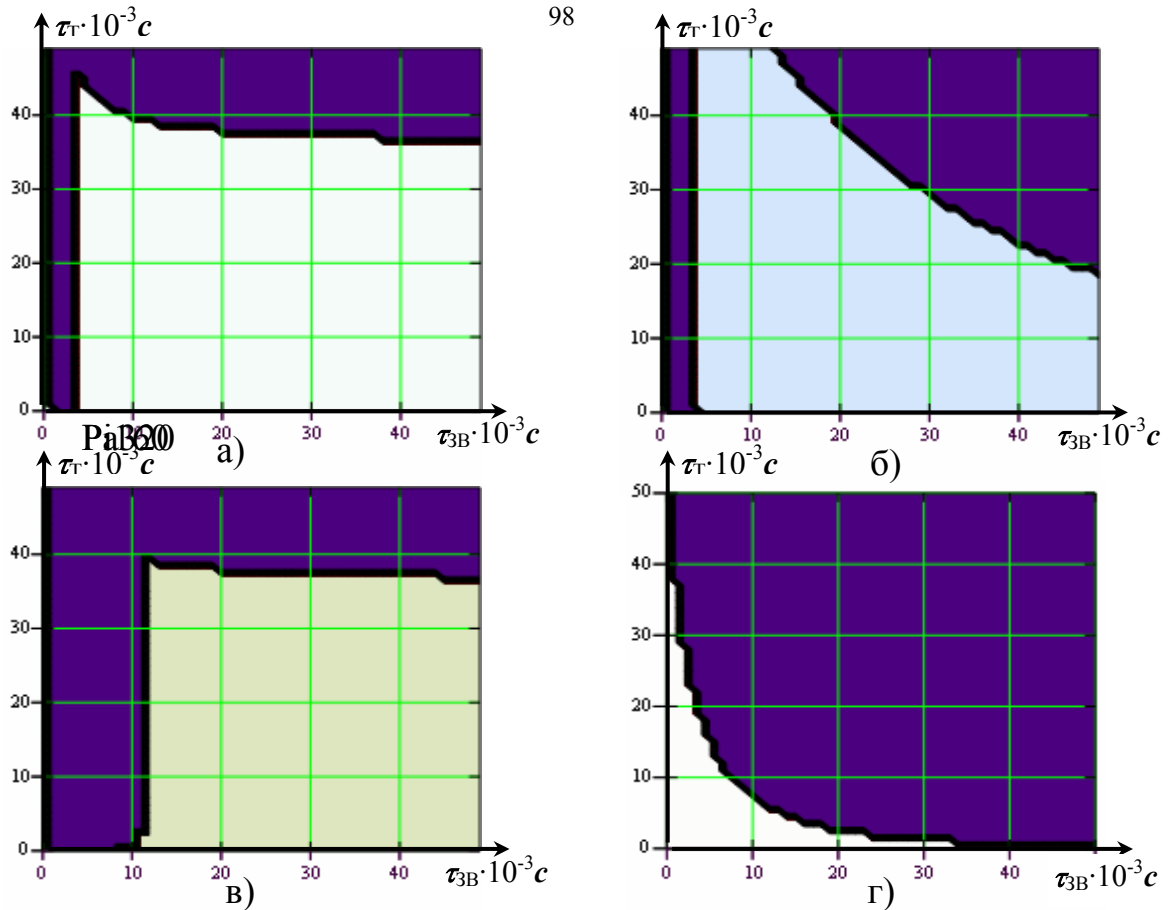


Рисунок 3.26 Зоны устойчивых колебаний (светлое) в зависимости от τ_r и τ_{3B} (а - $\tau_s = \tau_g = 40 \text{ мс}$, $F_p = 2,5 \text{ кГц}$; б - $\tau_s = \tau_g = 40 \text{ мс}$, $F_p = 4 \text{ кГц}$; в - $\tau_s = \tau_g = 30 \text{ мс}$, $F_v = 2,5 \text{ кГц}$; г - $\tau_s = \tau_g = 30 \text{ мс}$, $F_v = 4 \text{ кГц}$)

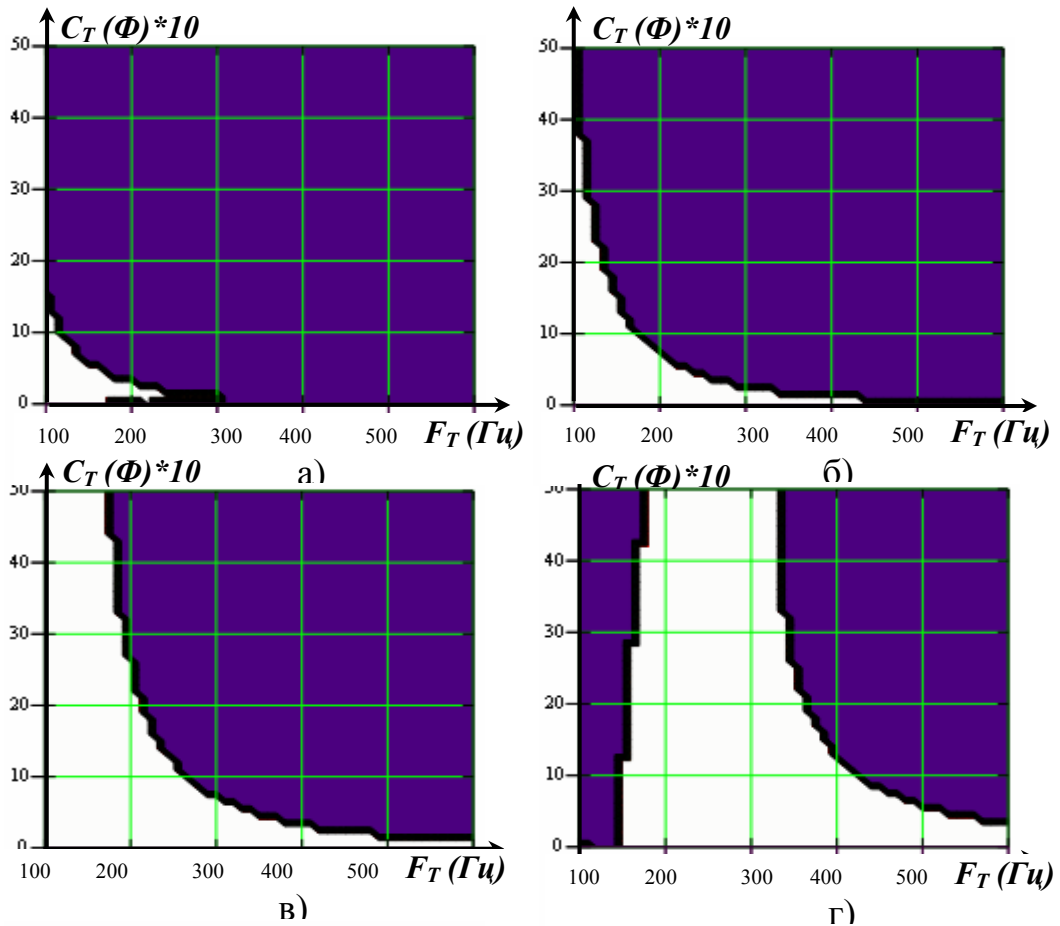


Рисунок 3.27 Зоны устойчивых колебаний (белое) в зависимости от F_T и C_T при $\tau_{3B} = 10^{-2} \text{ мс}$, (а - $\tau_n = 0,1 \text{ мс}$; б - $\tau_p = 0,5 \text{ мс}$; в - $\tau_p = 0,25 \text{ мс}$, г - $\tau_p = 1,25 \text{ с}$)

Выводы по главе:

Проведена оценка параметров моделей образования звучной речи. Определены их размерности, и установлена связь между акустическими и механическими величинами и соответствующими величинами электрических аналогов. Установлены возможные диапазоны изменения основных параметров моделей.

Выявлена роль гортанного желудочка. Показано, что существенное значение на формирование звучной речи оказывает собственная, резонансная частота гортанного желудочка.

Проведен анализ начальных условий и стационарных решений. На основе этого установлена связь между рядом параметров моделей образования звучной речи. В частности, установлены ограничения на совокупность параметров, характеризующих соотношениями между емкостью и индуктивностью звеньев речеобразующего тракта, величиной потерь в голосовой щели и коэффициентов, учитывающих влияние перепада напряжений на движения голосовых складок.

Установлены допустимые значения для постоянных времени трахеи и звеньев речеобразующего тракта, при которых обеспечивается близкое к реальному раскрытие голосовой щели. Установлены граничные условия для коэффициентов, учитывающих влияние сил Бернулли.

Исследованы области, в которых возникают устойчивые колебания в моделях образования нормальной и пищеводной звучной речи. На основе этого установлены допустимые значения для постоянных времени трахеи и звеньев речеобразующего тракта, а также постоянной времени раскрытой голосовой щели. Полученные данные хорошо коррелируются с известными экспериментальными данными других авторов [117, 125], что косвенно подтверждает адекватность моделей.

Установлено, что в полной речеобразующей системе при раскрытии голосовых складок возникают дополнительные резонансные частоты, которые можно рассматривать как известный механизм появления ложных формант.

ГЛАВА 4. ИССЛЕДОВАНИЕ СТРУКТУРЫ СИГНАЛОВ ЗВУЧНОЙ РЕЧИ

Исследование структуры сигналов звучной речи преследовало следующие цели:

- выявление закономерностей построения звучной нормальной и пищеводной речи;
- анализ структуры сгенерированных в моделях сигналов нормальной и пищеводной речи и сравнение их с нормальной речью;
- субъективная оценка качества сгенерированных в модели сигналов нормальной звучной речи.

Решение этих вопросов позволяет:

- установить адекватность разработанных моделей нормальной и пищеводной звучной речи;
- провести оценку качества реабилитации прооперированных пациентов;
- дать рекомендации по модификации методики обучения пищеводной звучной речи.

При проведении этих исследований использовался программный комплекс [10, 15, 19, 20], непосредственное и основное участие в разработке которого принимал Пономарев А.А.

4.1 Комплекс для исследования структуры речевых сигналов

Данный исследовательский комплекс содержит набор алгоритмов и предоставляет пользовательский интерфейс, позволяющий:

- формировать изображения типа «видимая речь»;
- выводить в различных масштабах выбранные участки речевого сигнала на всех стадиях обработки;
- проводить детальный анализ спектральных компонентов речевого сигнала.

Описываемое исследовательское средство постоянно развивается и дополняется по мере разработки новых алгоритмов и модификации старых. В данный момент он содержит ряд программных средств, которые включают:

- цифровую фильтрацию речевого сигнала;
- сегментацию сигнала по параметрам и по спектру;
- выделение частоты и периода основного тона;
- выделения формантных частот.

Кроме того, отладку и предварительные испытания вновь создаваемых алгоритмов удобнее вести в стандартных математических прикладных пакетах. Для этого исследовательский комплекс имеет возможность связи с математическими приложениями (MathCAD, MathLab) на файловом уровне. Это позволяет быстро моделировать новые алгоритмы обработки и анализировать результаты их работы в исследовательской среде MathCAD и MathLab. Единственное, что модели сформированные в MathCAD позволяют рассчитать синтезированный звук длительностью 0,3 с за 20 мин машинного времени на IBM PC класса AthlonXP 2400+ с 512 Мб ОЗУ, тогда как программа написанная на языке Фортран [131] производит расчет за одну две минуты на этой же машине. Описываемое инструментальное средство для исследования структуры речевых сигналов реализовано в среде Windows и может использоваться на любом IBM-совместимом компьютере с операционной системой Windows 95 и выше.

Особенностью программного комплекса является следующее:

- спектральный анализ («видимая речь») производится нерекурсивными цифровыми фильтрами таким образом, что ось частот имеет логарифмический масштаб. Это позволяет повысить точность измерения частотных компонентов в низких частотах, т.е. в области частот основного тона;
- окно анализа может устанавливаться пользователем. Обычно оно устанавливается разным по всему частотному диапазону. Это позволяет при исследованиях гибко выбирать разрешающую способность по частоте и времени.

Такие требования к обработке речевых сигналов были оформлены на основе экспериментальных данных на психоакустике восприятия [124,133]. К сожалению, в работе только частично были учтены все фонетические особенности речи, описание которых можно найти подробно в [9, 21, 22, 23, 42, 43, 44, 45, 46, 47, 62, 80, 99, 100, 101, 109, 119, 144, 153, 154, 163, 172, 173].

Основные соотношения для обработки речевых сигналов представлены в табл. 4.1. Импульсные переходные характеристик системы фильтров имеют вид [145]

$$h_c(t, x) = \frac{\omega_0 \cdot e^{(\beta-1)x}}{2,4 \cdot \sqrt{\pi} Q_0} \cdot \exp \left[-\frac{\omega_0^2 \cdot e^{2(\beta-1)x} \cdot t^2}{5,76 \cdot Q_0^2} \right] \cdot \cos(\omega_0 \cdot e^{-x} \cdot t), \quad (4.1)$$

$$h_s(t, x) = \frac{\omega_0 \cdot e^{(\beta-1)x}}{2,4 \cdot \sqrt{\pi} Q_0} \cdot \exp \left[-\frac{\omega_0^2 \cdot e^{2(\beta-1)x} \cdot t^2}{5,76 \cdot Q_0^2} \right] \cdot \sin(\omega_0 \cdot e^{-x} \cdot t), \quad (4.2)$$

где ω_0 – верхняя частота анализа;

Q_0 – добротность фильтра на частоте ω_0 ;

$$x_{\max} = \ln \left(\frac{\omega_0}{\omega_H} \right);$$

ω_H – нижняя частота анализа;

$$Q_0 = Q_0 \cdot e^{-x_{\max}};$$

x – пространственная координата;

$$\omega_p(x) = \omega_0 \cdot e^{-x} \text{ – резонансная частота фильтра с координатой } x;$$

При цифровой обработке используется:

$$t = i \cdot T_k; \quad x = k \cdot \Delta x,$$

где T_k – шаг квантования по времени (обычно $T_k = \frac{1}{22050}$);

Δx – шаг квантования по координате x .

Уже много лет во всем мире делаются попытки создания универсальных систем распознавания и синтеза речи. Но до сих пор наилучший результат в области синтеза достигается при использовании методов компиляции [74, 149, 150], то есть речевой сигнал составляется из заранее подготовленных естественных сигналов. Для распознавания также используется предварительное составление словаря образцов из наговоренных слов и фраз.

Система распознавания или синтеза не может быть универсальной, если она заранее ориентирована на конкретного диктора, а существующие системы всегда ориентированы на конкретного пользователя [33, 34, 49, 57, 64, 66, 67, 69, 120]. Универсальный анализатор или синтезатор речи не должен быть привязан к индивидуальным особенностям человека, а должен обладать знанием общих для всех людей правил построения речи и речевого сигнала. Эти правила, несомненно, существуют, иначе бы человеку требовалось немало времени, чтобы начать понимать собеседника.

Однако на сегодняшний день раскрыты далеко не все правила построения речевого сигнала. Поэтому задачи анализа структуры речевого сигнала, ориентирование на выделение параметров, характеризующих конкретного диктора, фонемный состав и т.д., по-прежнему остаются актуальными. Проведение данных работ требует наличия соответствующих средств автоматизации экспериментальных исследований. В данной главе рассматривается структура инструментальных средств автоматизированного комплекса для исследования речевого сигнала, а также результаты экспериментальных исследований для нормальной и фарингеальной речи.

Инструментальные средства ориентированы на исследовательские алгоритмы обработки звука как в области распознавания, так и в области синтеза речи. При анализе и синтезе речи инструментальные средства позволяют использовать уже известные алгоритмы и механизмы речеобразования и речевосприятия.

Основные соотношения математического обеспечения разработанного исследовательского комплекса сведены в табл. 4.1. В соответствии с ними разработан программный комплекс, приведенный на рис. 4.1.

На основе результатов обработки в соответствии с табл. 4.1 по мощности спектральных компонент производится выделение спектральных максимумов и анализ временной структуры составляющих $y_c(I,k)$, $y_s(I,k)$. На основе их производится синхронизация с периодом основного тона, выделение отрезков речевого сигнала на периоде основного тона, синхронный спектральный анализ речевого сигнала на периоде основного тона.

На основе мгновенной частоты, измеренной в области спектрального максимума, определяются частоты формант, частота основного тона.

Таблица 4.1. Основные соотношения

Функция	Представление
Цифровая фильтрация [145]	$y_c(i, k) = \sum_{j=i-n}^{i+n} x_j \cdot h_c(j, k),$ $y_s(i, k) = \sum_{j=i-n}^{i+n} x_j \cdot h_s(j, k),$ <p>где y_c, y_s – сигнал после фильтрации; x_j – исходный речевой сигнал; $h_c(j, k), h_s(j, k)$ – импульсные переходные характеристики системы фильтров с косинусной и синусной составляющими;</p>
Мощности спектральных компонентов	$M(i, k) = y_c^2(i, k) + y_s^2(i, k),$ <p>где y_c^2, y_s^2 – сигналы с выхода системы цифровых фильтров синусной и косинусной составляющих.</p>
Мгновенная частота	$w_k(t) = \frac{y_s(k, t) \cdot y_c(k, t-1) - y_s(k, t-1) \cdot y_c(k, t)}{T_k \cdot M(k, t)},$ <p>где t – дискретное время; T_k – шаг квантования; $M(k, t)$ – мощности спектральных компонентов.</p>
Одновременная и последовательная маскировка	$F(i, k) = \text{if}(M(i, k) - P_0(i, k) > 0, 1, 0),$ $T(i, k) = \text{if}(M(i, k) - P_p(i, k) > 0, 1, 0)$ <p>где $P_0(i, k)$ и $P_p(i, k)$ – функции одновременной и последовательной маскировки</p>

Основная задача использования описываемого инструментального средства – детальное исследование структуры речевого сигнала, а именно – мгновенных спектров. Анализ входного сигнала ведется по множеству каналов, что отражено на рис. 4.1, 4.2. На этих рисунках отражены амплитудно-частотные картины, где по оси X отложено дискретное время (выборки), по оси Y – частота (номер канала анализа), а цветом отображается амплитуда сигнала.

На рисунках представлены фрагменты косинусных компонентов амплитудно-частотных картин переходов звука «Л» в звук «А» в слове «ЛА» для мужского (рис. 4.2) и женского (рис. 4.3) голосов. На рис. 4.4, 4.5

приведены динамические спектрограммы интенсивности для того же слога тех же голосов соответственно. Снизу амплитудно-частотных картин выведен сигнал на выходе 36-го фильтра системы цифровых фильтров для мужского голоса и 32-го фильтра для женского, а справа от них – распределение амплитуд по каналам для дискретного момента времени 3297 и 3480 соответственно.

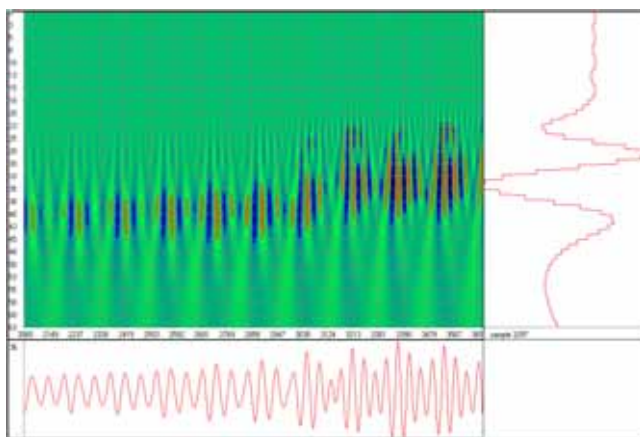


Рисунок 4.2

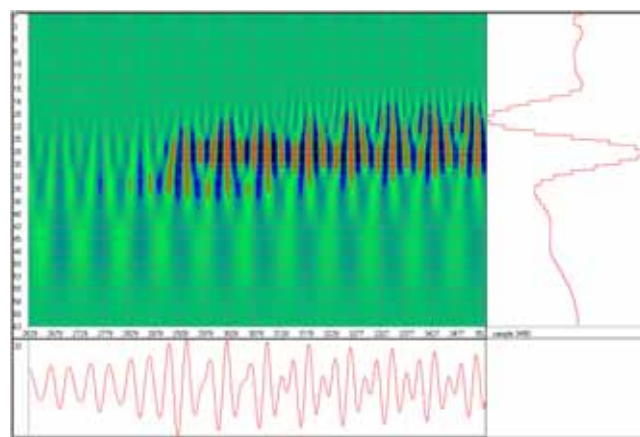


Рисунок 4.3

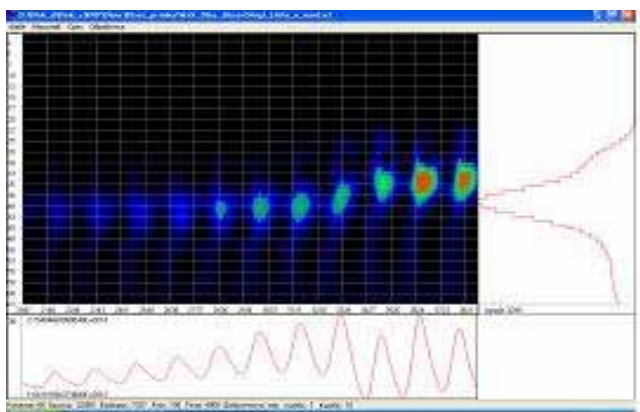


Рисунок 4.4

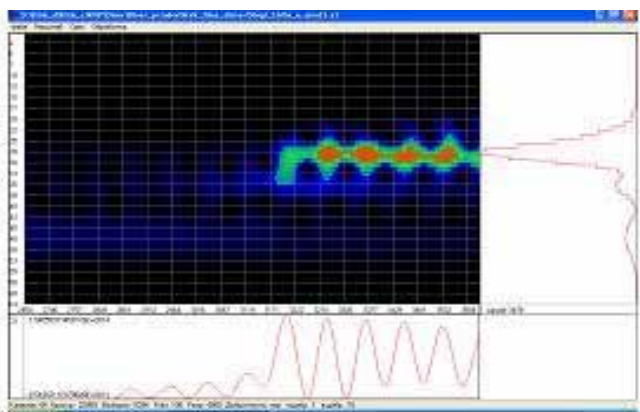


Рисунок 4.5

Дальнейшая работа по нахождению характеристик и параметров речевого сигнала ведется в частотной области. Описываемый исследовательский комплекс позволяет строить графики зависимости изменения частоты основного тона от дискретного времени.

Кроме того, использование механизмов маскировки позволяет выявить структурные особенности речевого сигнала. Маскирующие функции $P_0(t, x)$ и $P_P(t, x)$ определяется как свертка мощности спектральных компонентов с весовыми функциями, представляемые на основе соотношений (4.1) и (4.2) [95].

4.2 Исследование влияния разности давлений в трахее и ротовой полости на частоту основного тона

На первом этапе исследовался характер изменения частоты основного тона. В частности, при произнесении слогов типа звонкий взрывной-гласный следует отметить, что смычка на звонком взрывном повышает давление внутри ротовой полости. Тогда если предположение о том, что частота основного тона определяется разностью давлений в трахее и ротовой полости верно, то частота основного тона на участке взрывного звука должна быть меньше. Кроме того, следует ожидать, что для слогов типа «БИ» несмотря на то, что сечение речеобразующего тракта для звука «И» весьма мало, но нет смычки, частота основного тона должна быть выше, чем у звонкой согласной «Б». На рис. 4.8, 4.9, 4.10 и 4.11 приведены результаты этих исследований. На рис. 4.6 и 4.7 приведены исходный речевой сигнал и частота основного тона, на остальных - только частота основного тона.

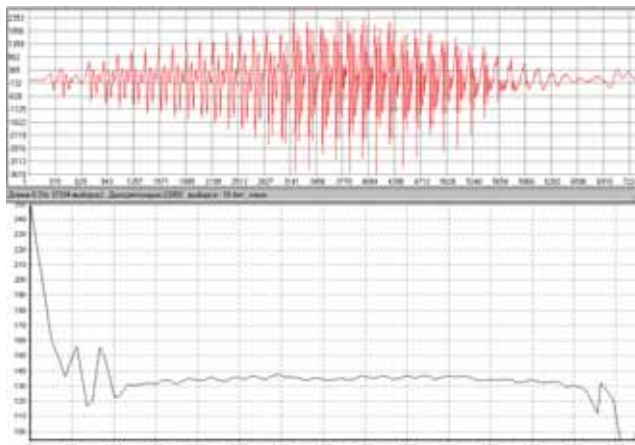


Рисунок 4.6

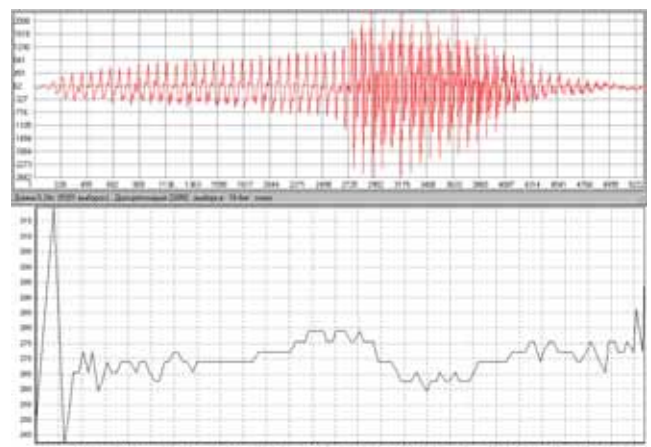


Рисунок 4.7

Данные зависимости изменения частоты основного тона от дискретного времени приведены для двух дикторов. Рис. 4.6, 4.7 – для слога «ЛА», рис. 4.8, 4.9, 4.10, 4.11 – для слога «БИ», рис. 4.12, 4.15 – для слога «УИ» а рисунки 4.16, 4.17 – для слога «ИУ» сказанных мужским и женским голосом соответственно.

При произношении слога «ЛА» частота основного тона меняется плавно, и по данным характеристикам достаточно затруднительно сказать, где заканчивается одна фонема и начинается другая. Это можно объяснить тем, что при произнесении звука «Л» имеется достаточная площадь излучения между

языком и зубами и часть воздушного потока проходит через ротовую полость, т.е. давление в ротовой полости почти не повышается. Абсолютно противоположное можно сказать про слог «БИ». На представленных характеристиках хорошо заметен характерный скачок частоты при переходе с фонемы «Б» на фонему «И». Характерны графики изменения частоты основного тона для слогов «УИ» и «ИУ». Вероятнее всего, изменение частоты основного тона определяется изменением избыточного давления в легких и трахее.

Как следует из этих данных, можно с достаточной степенью уверенности заключить, что частота основного тона в значительной степени определяется разностью давлений в трахее и ротовой полости.

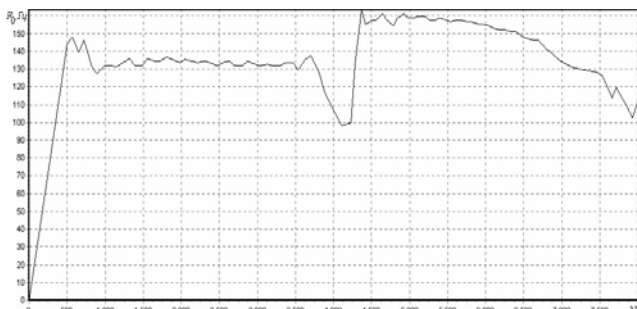


Рисунок 4.8 Слог «БИ», диктор
мужчина (20.09.03)

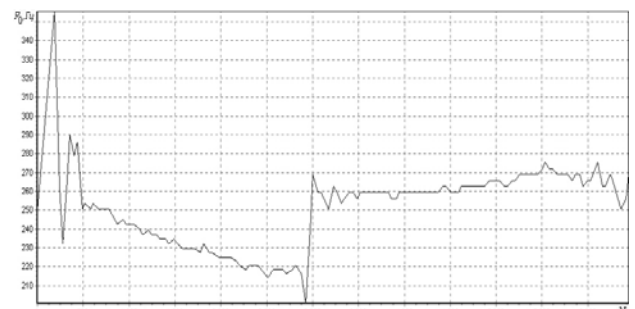


Рисунок 4.9 Слог «БИ», диктор
женщина (20.09.03)

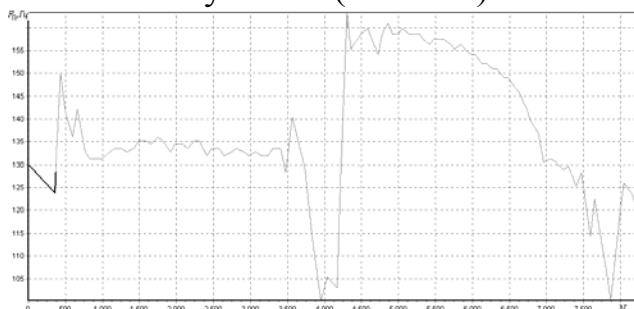


Рисунок 4.10 Слог «БИ», диктор
мужчина (22.09.03)

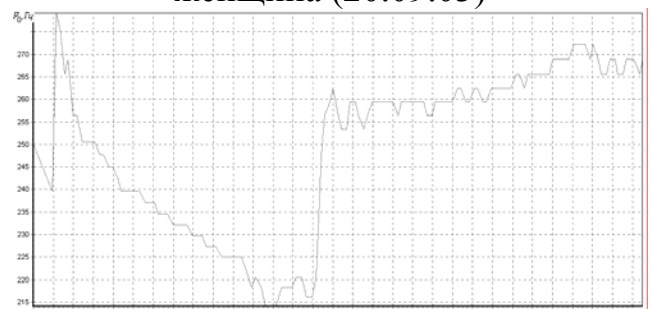


Рисунок 4.11 Слог «БИ», диктор
женщина (22.09.03)

Это позволяет сделать вывод, что основную роль в формировании частоты основного тона играет разность давлений. В этом случае нервные возбуждения на голосовых складках, вероятно, определяют их натяжение, т.е. собственную резонансную частоту. Это обстоятельство учтено в моделях частотой ω_c и табл. 2.7 и 2.8. Поэтому при обучении пищеводному голосу необходимо сформировать условный рефлекс для управления псевдоголосовой

щелью. Так как он формируется в достаточно зрелом возрасте, то эта обратная связь, по-видимому, неустойчива, и следует ожидать, что частота основного тона для пищеводного голоса будет нестабильной.

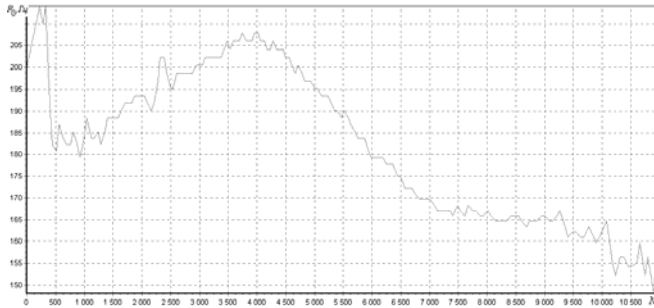


Рисунок 4.12 Слог «УИ», диктор мужчина

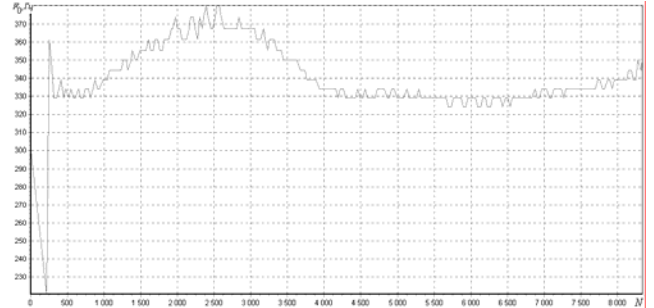


Рисунок 4.13 Слог «УИ», диктор женщина

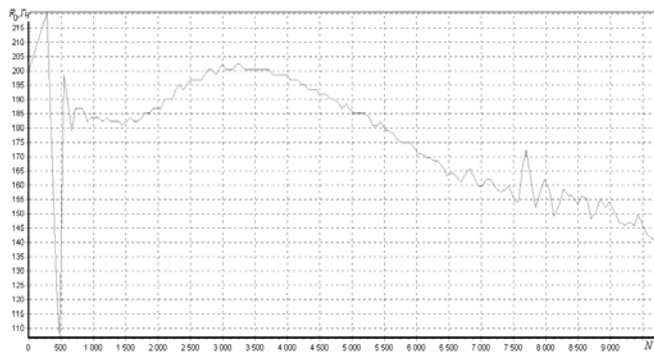


Рисунок 4.14 Слог «УИ», диктор мужчина

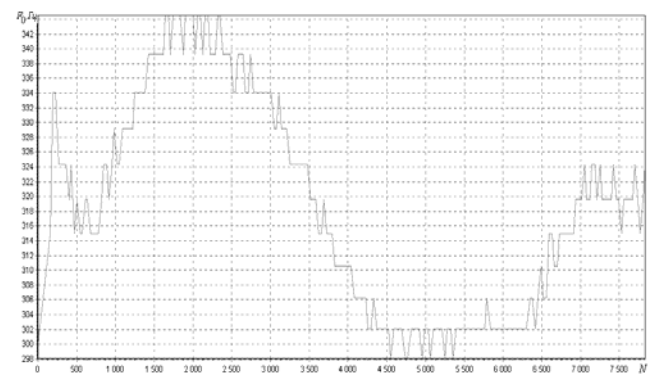


Рисунок 4.15 Слог «УИ», диктор женщина

Эти данные важны также при распознавании речевого сигнала. Во-первых, по положению этого скачка можно достаточно точно разграничить сегменты в речевом потоке. Во-вторых, по такому характерному перепаду можно с достаточно большой вероятностью охарактеризовать предшествующий ему сегмент речи как сегмент, содержащий взрывной звук, существенно сократив набор возможных фонем, которые могут иметь место в данном сегменте.

4.3 Синтез речевых сигналов гласных звуков

Синтез речевых сигналов гласных звуков проводился на основе моделей, представленных во второй главе диссертации. Предварительно был проведен синтез речевых сигналов при принудительном управлении движением голосовыми складками, т.е. в выражении (2.17) полагалось, что $B=0$. В этом случае полностью исключалось влияние разности давлений в трахее и ротовой полости. Такой режим генерации речевых сигналов соответствует нейрохроноксической гипотезе колебания голосовых складок.

Субъективная оценка качества речевых сигналов показала, что они имеют низкую естественность [10, 15, 56, 142]. Это позволяет утверждать, с учетом косвенных данных по измерению частоты основного тона для слогов типа «БИ», что нейрохроноксическая гипотеза не соответствует возможным механизмам фонации. Этот вывод важен при обучении пищеводной звучной речи.

Последующие эксперименты по синтезу речевых сигналов проводились для $B=1$, но считалось, что в (2.17) $F_H(t)=0$. Таким образом, фонация полностью определялась разностью давления в трахее и легких. Параметры моделей речеобразующей системы выбирались из условий образования устойчивых колебаний.

На рис. 4.16 – 4.31 приведены результаты синтеза речевых сигналов гласных звуков. Субъективно данные речевые сигналы являлись вполне естественными. Это дает основание полагать, что предложенная модель движения голосовых складок наиболее полно соответствует реальным механизмам фонации.

Данный вывод позволяет считать, что и для пищеводной речи механизм фонации аналогичен.

На рис. 4.32 – 4.37 приведены результаты синтеза речевых сигналов пищеводной речи, полученные на основе модели, описанной во второй главе.

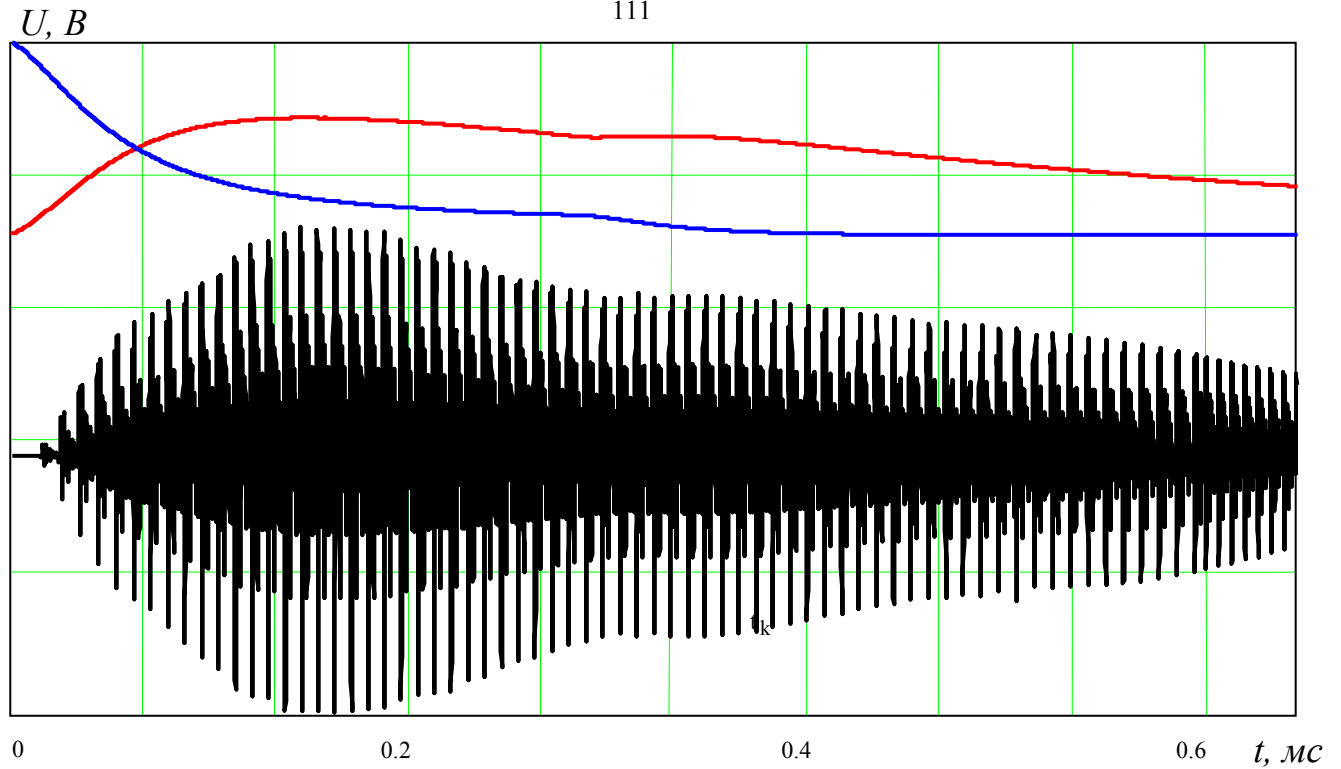


Рисунок 4.16 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «А»

$U, I, B, \text{мА}$

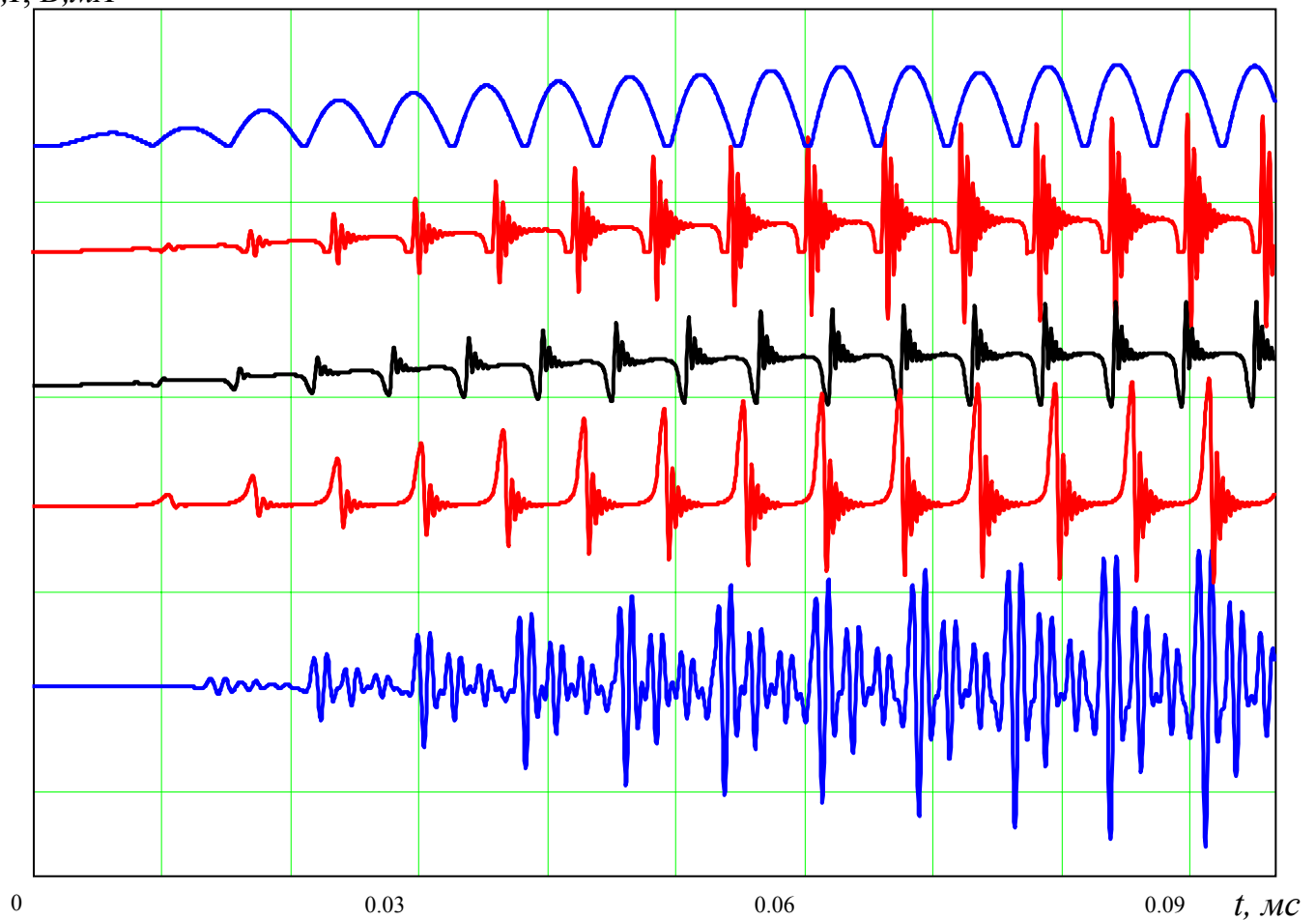


Рисунок 4.17 Формы сигналов на различных участках речеобразующего тракта для звука «А». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

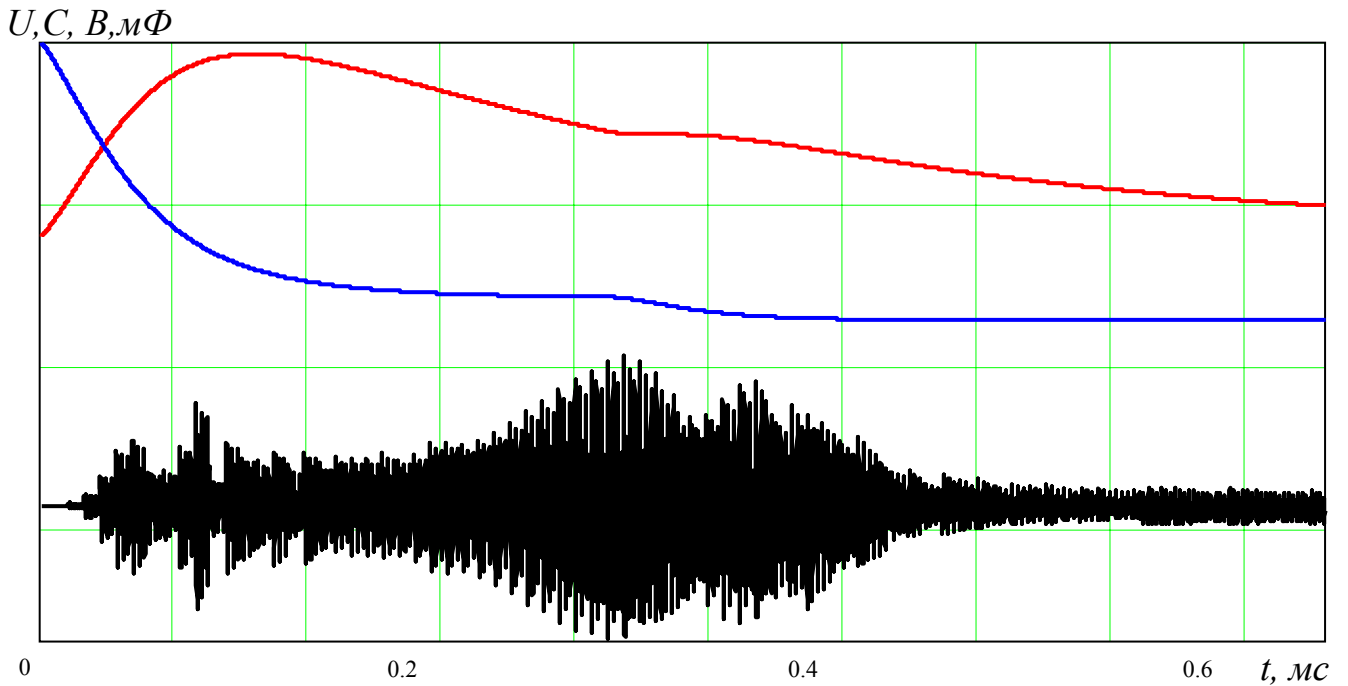


Рисунок 4.18 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «А» «мужской голос»
 $U, I, B, \text{мА}$

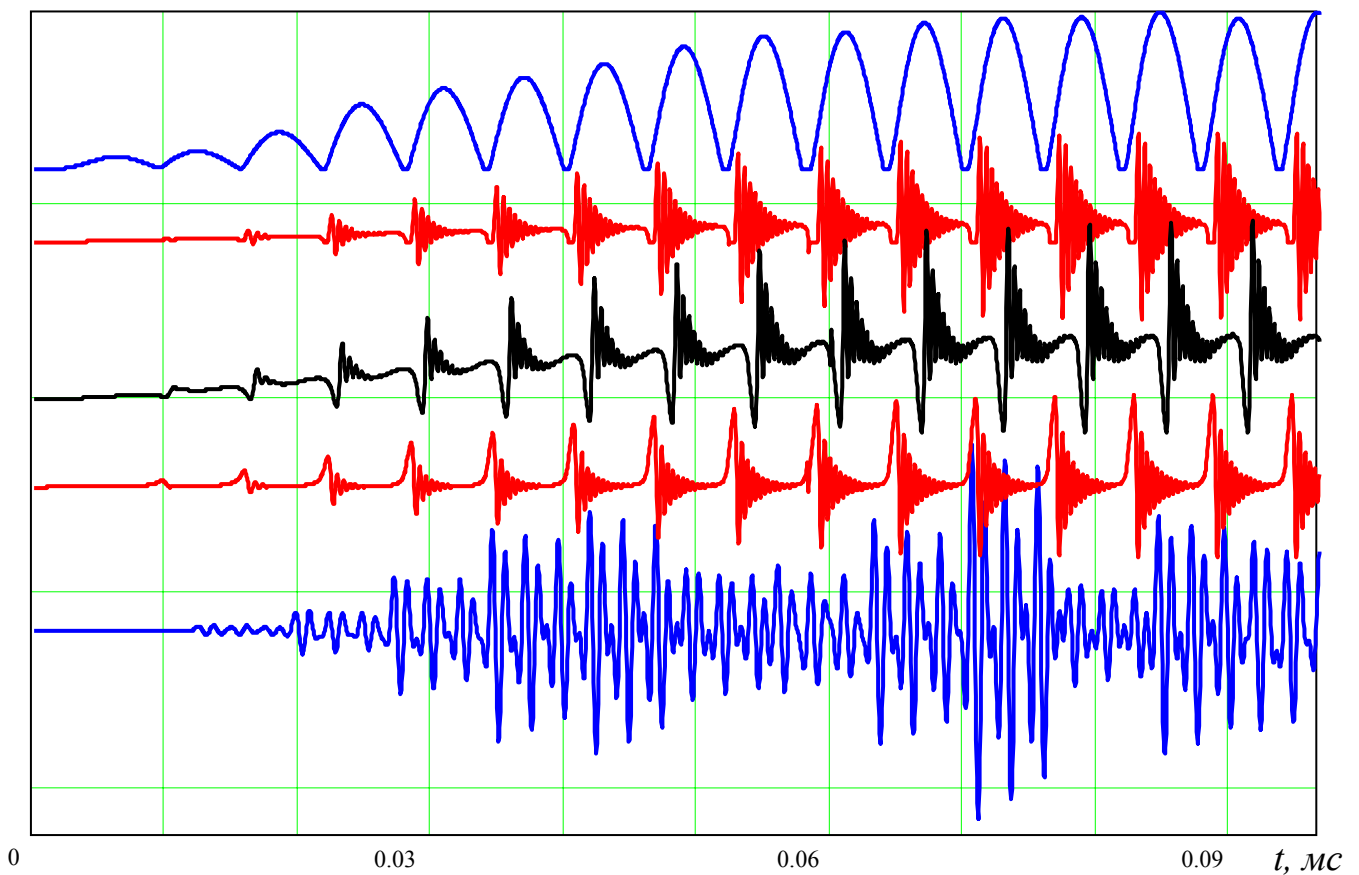


Рисунок 4.19 Формы сигналов на различных участках речеобразующего тракта для звука «А» «мужской голос». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта.

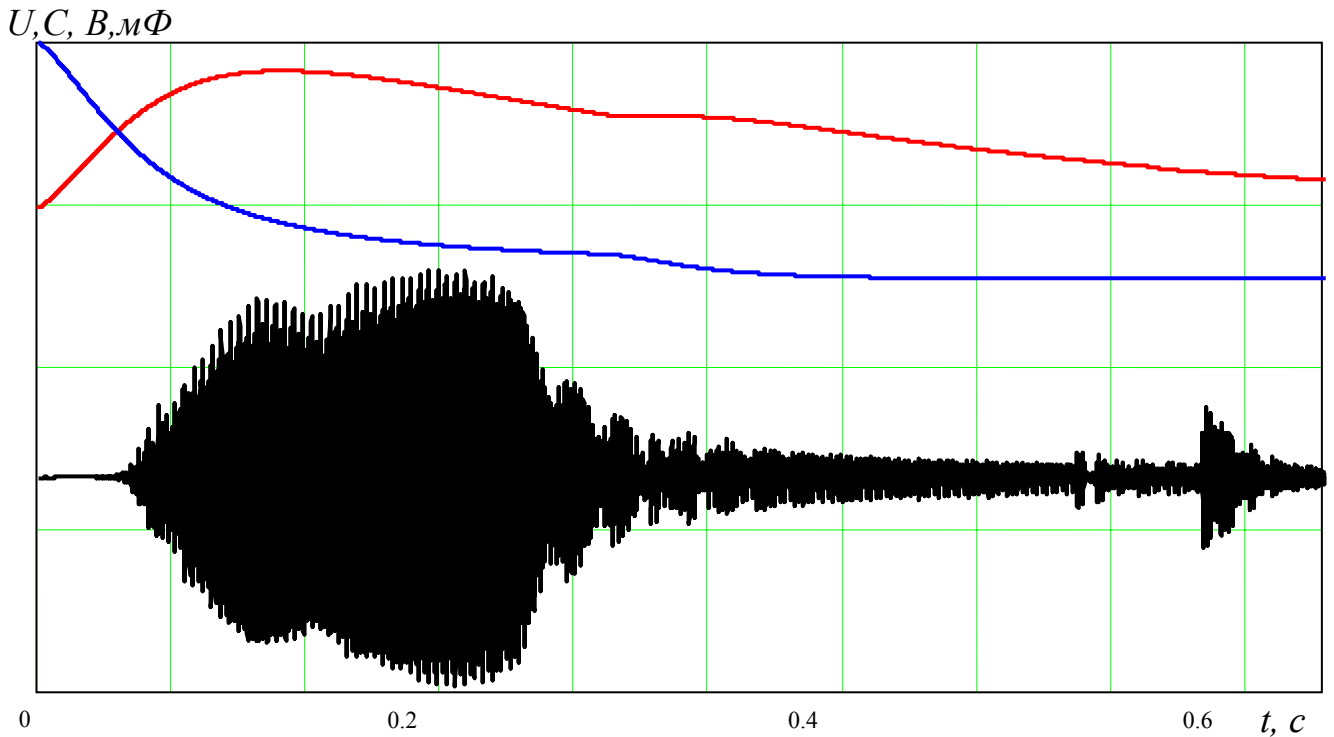


Рисунок 4.20 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «А» «женский голос»

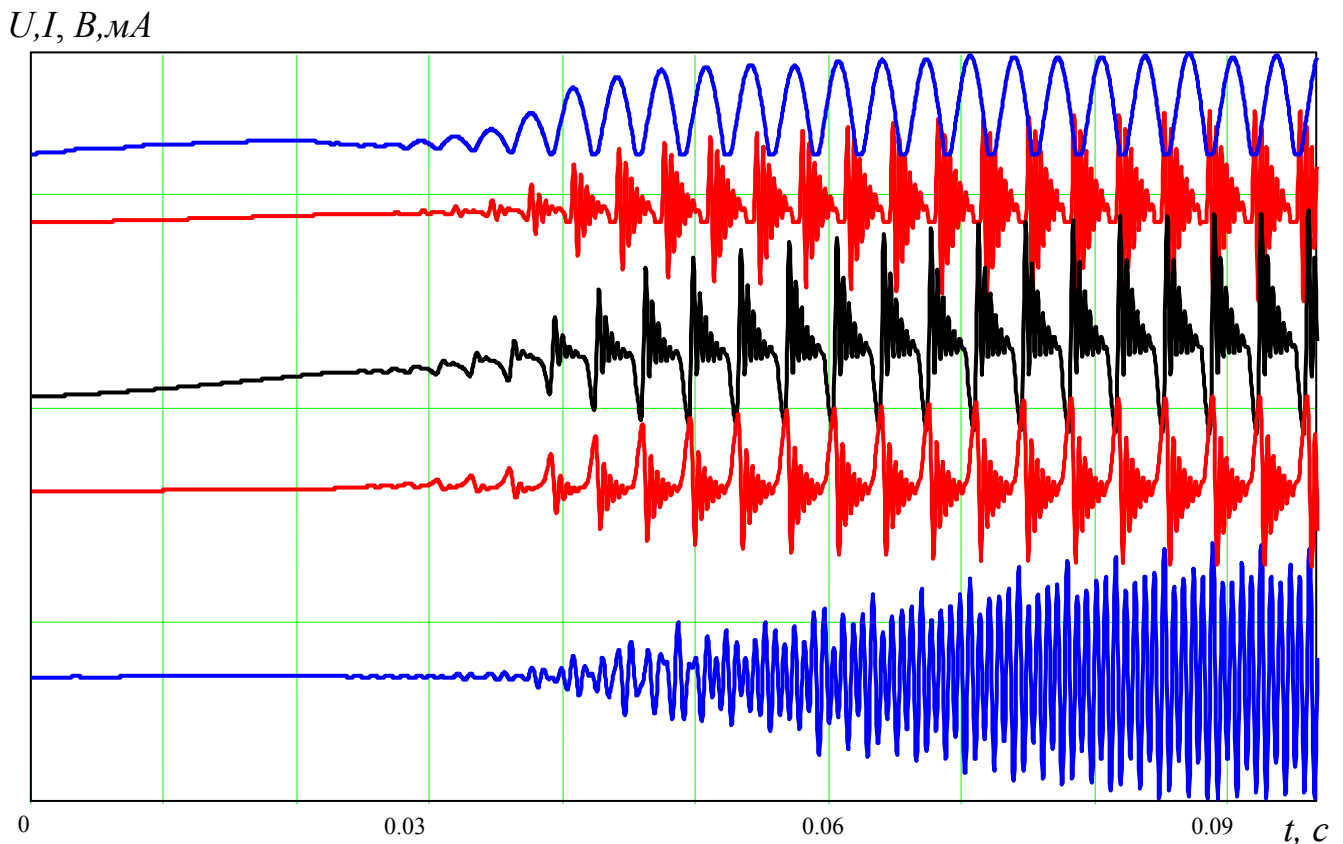


Рисунок 4.21 Формы сигналов на различных участках речеобразующего тракта для звука «А» «женский голос». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного) желудочке, разность давлений (напряжение на выходе речеобразующего тракта)

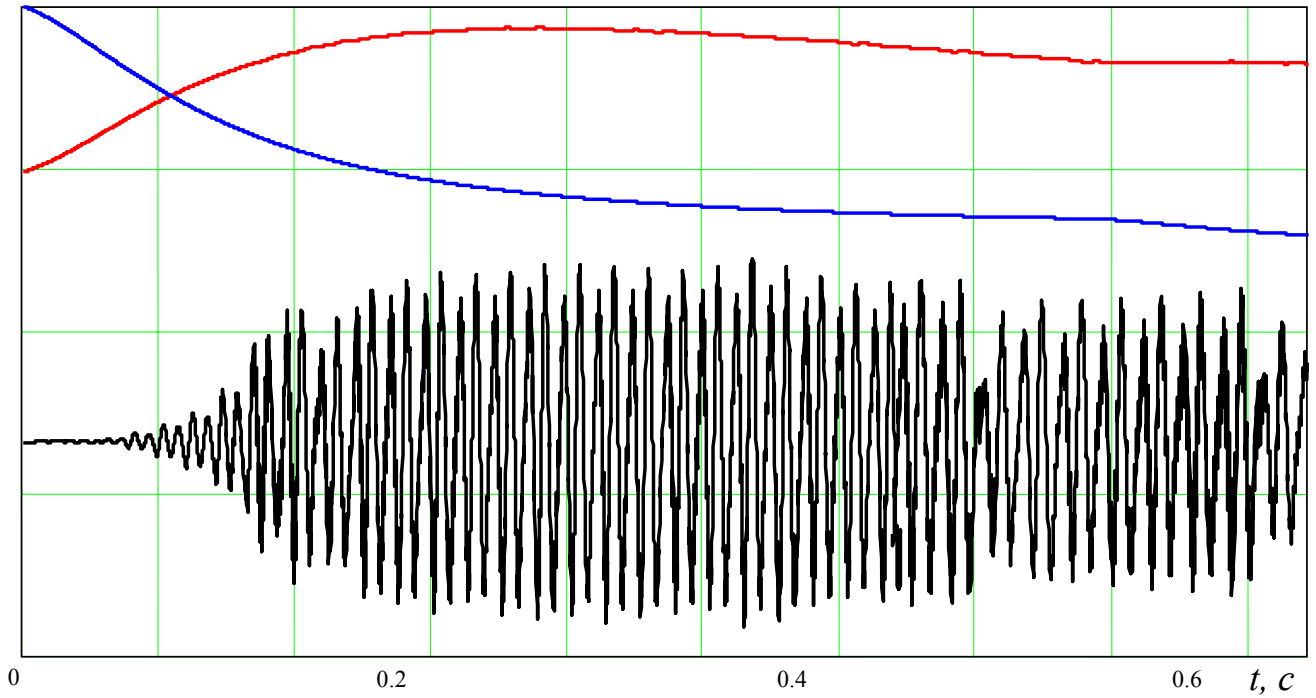
$U, C, B, \text{мФ}$ 

Рисунок 4.22 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «И»

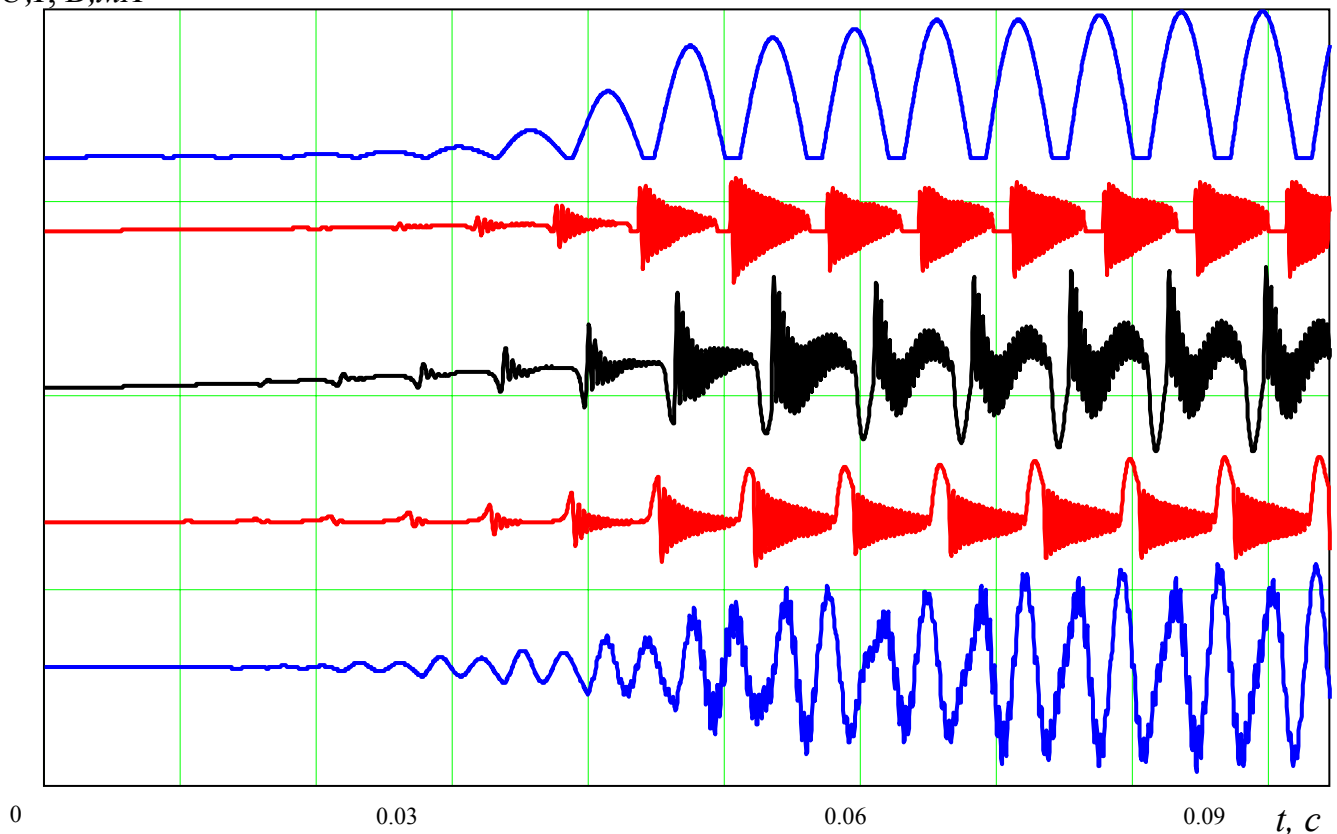
 $U, I, B, \text{мА}$ 

Рисунок 4.23 Формы сигналов на различных участках речеобразующего тракта для звука «И». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

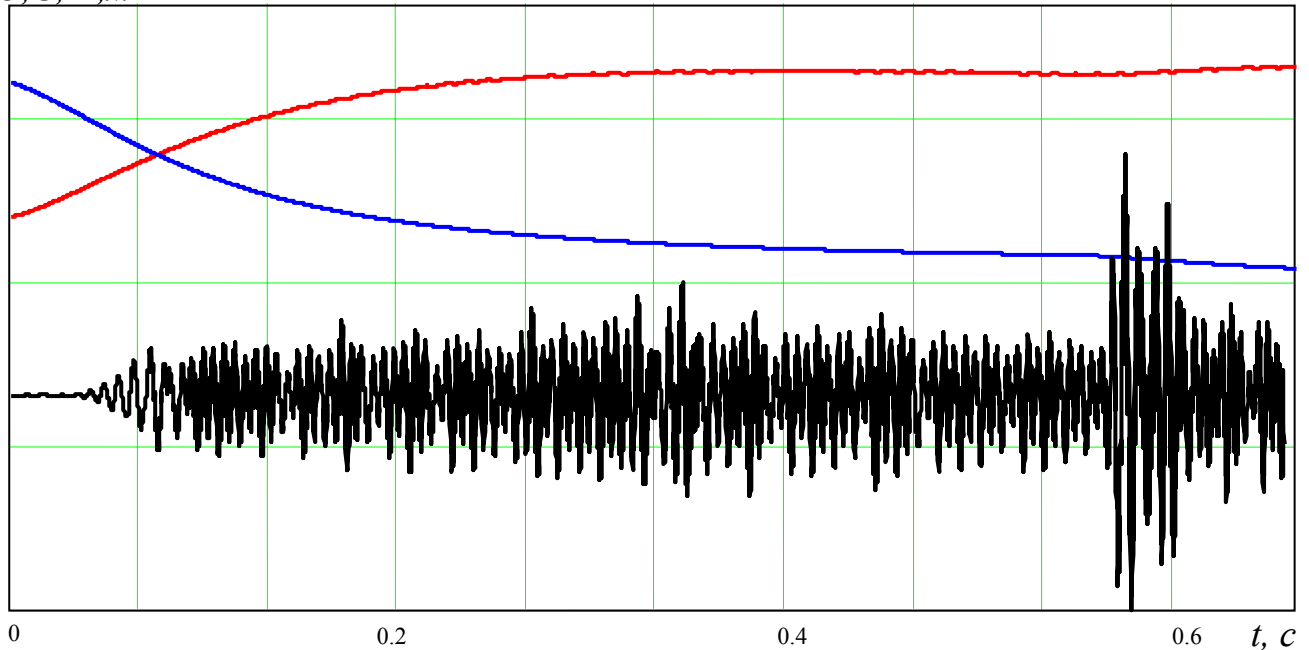
$U, C, B, \text{мФ}$ 

Рисунок 4.24 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «И»

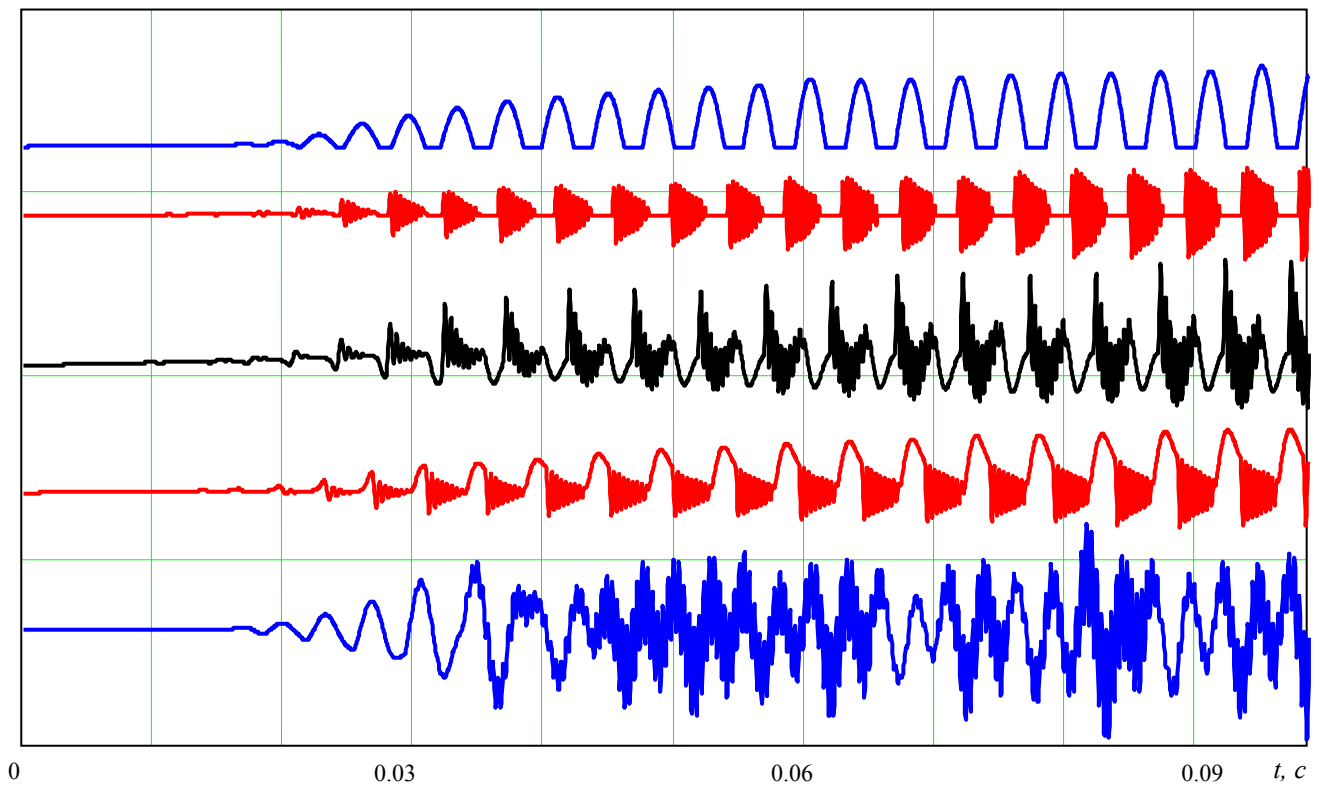
 $U, I, B, \text{мА}$ 

Рисунок 4.25 Формы сигналов на различных участках речеобразующего тракта для звука «И». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (над-складочного и подскладочного), напряжение на выходе речеобразующего тракта.

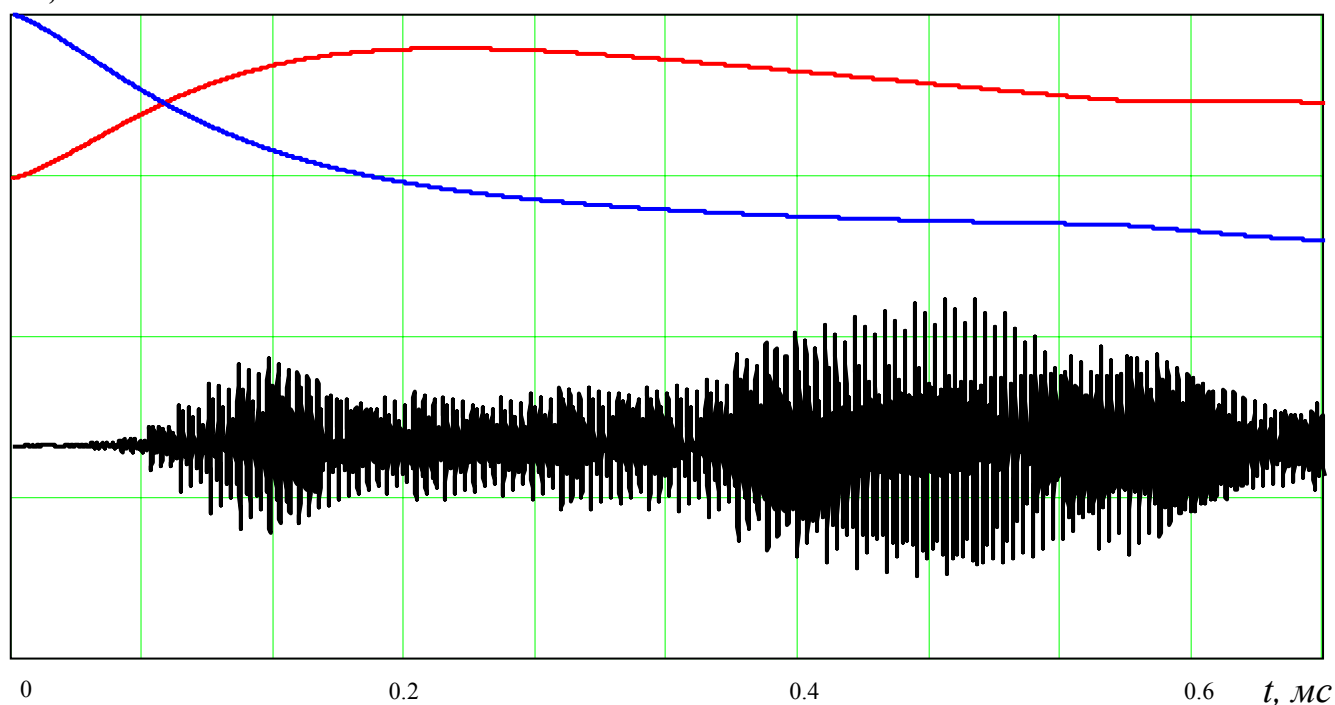


Рисунок 4.26 Избыточное давление в легких (---), Емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «О»

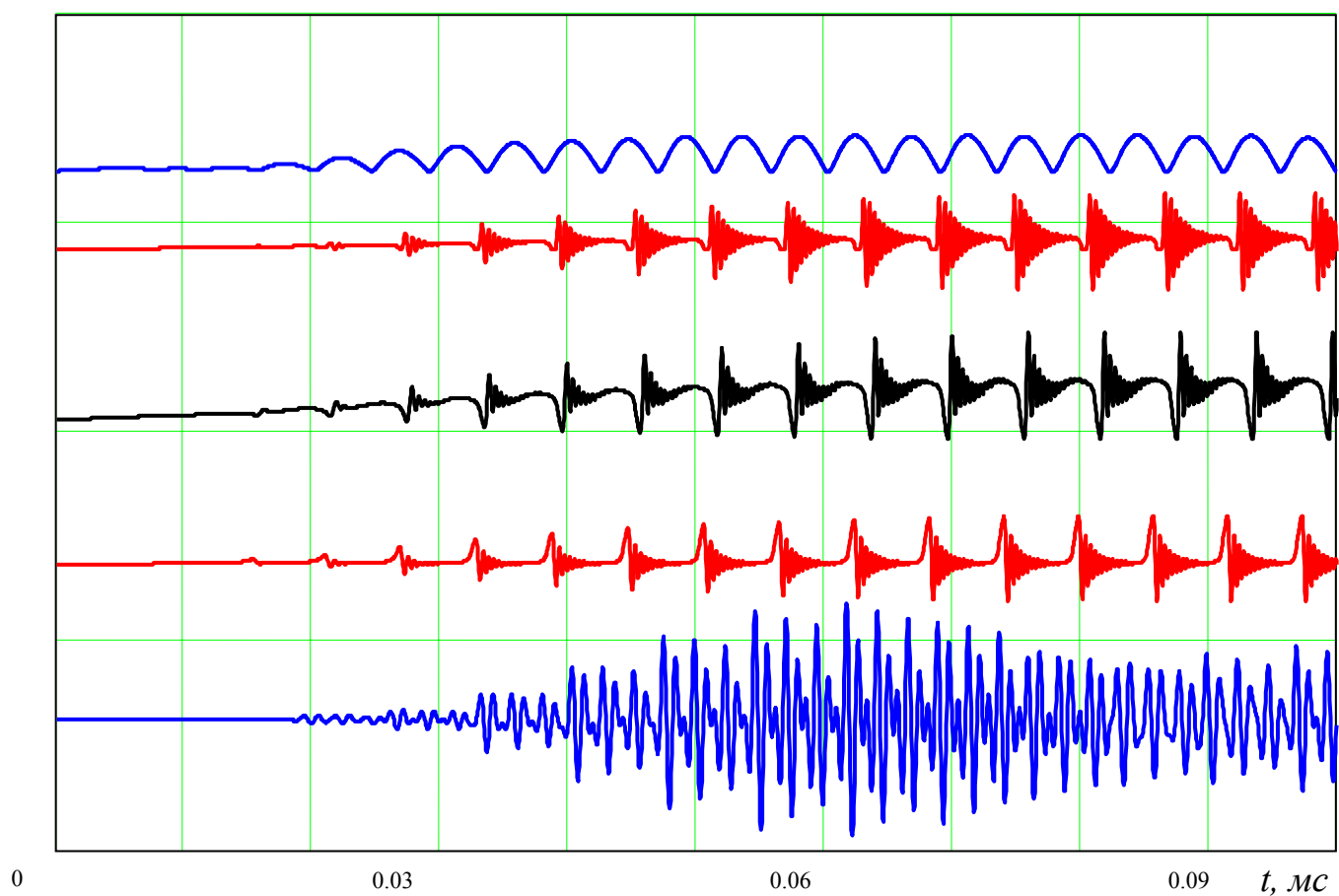
 U, I, B, mA


Рисунок 4.27 Формы сигналов на различных участках речеобразующего тракта для звука «О». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

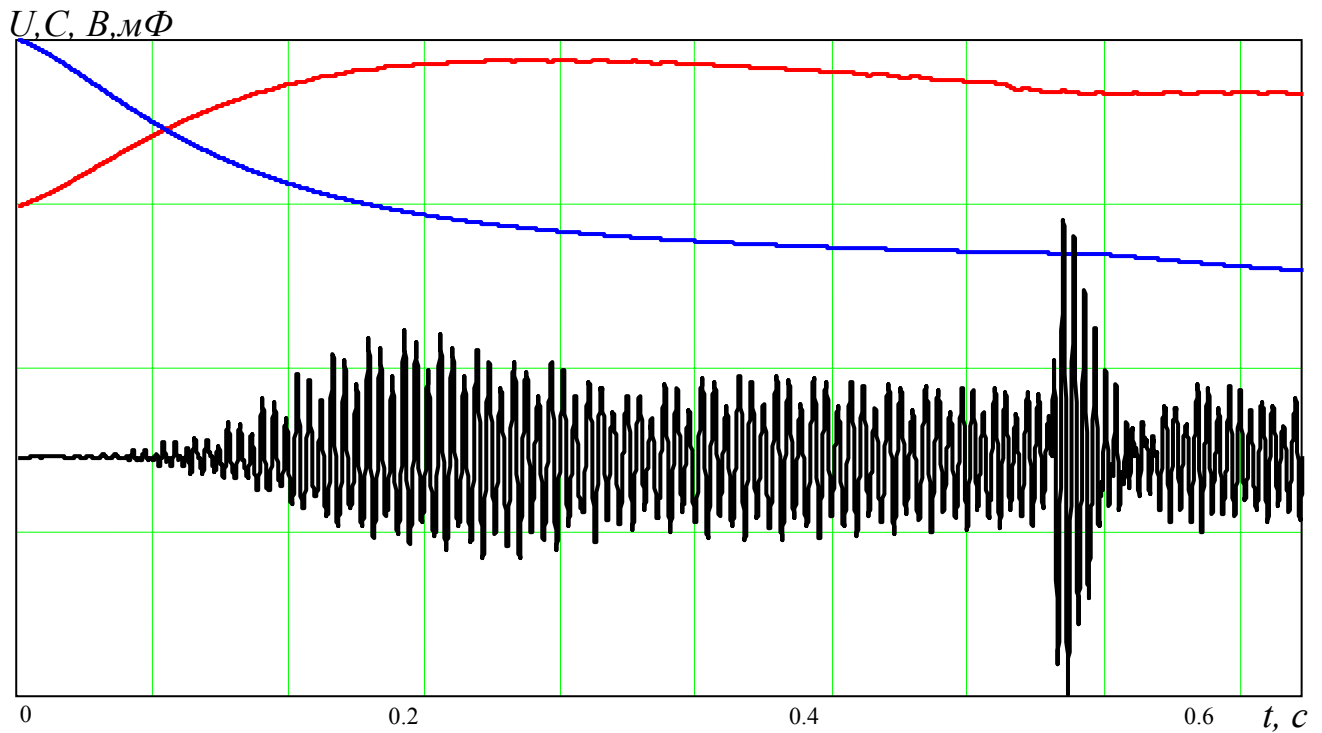


Рисунок 4.28 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «У»
 $U, I, B, \text{мА}$

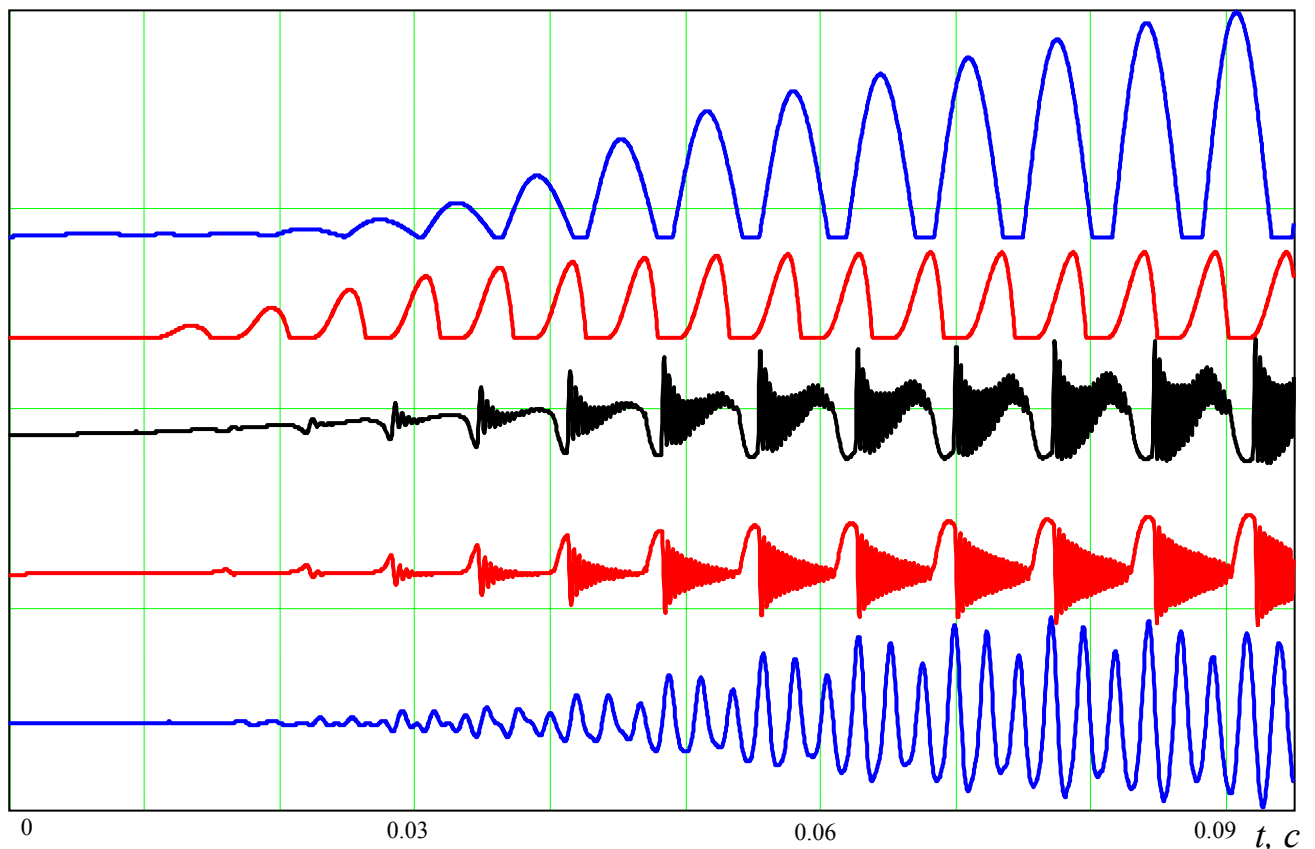


Рисунок 4.29 Формы сигналов на различных участках речеобразующего тракта для звука «У». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

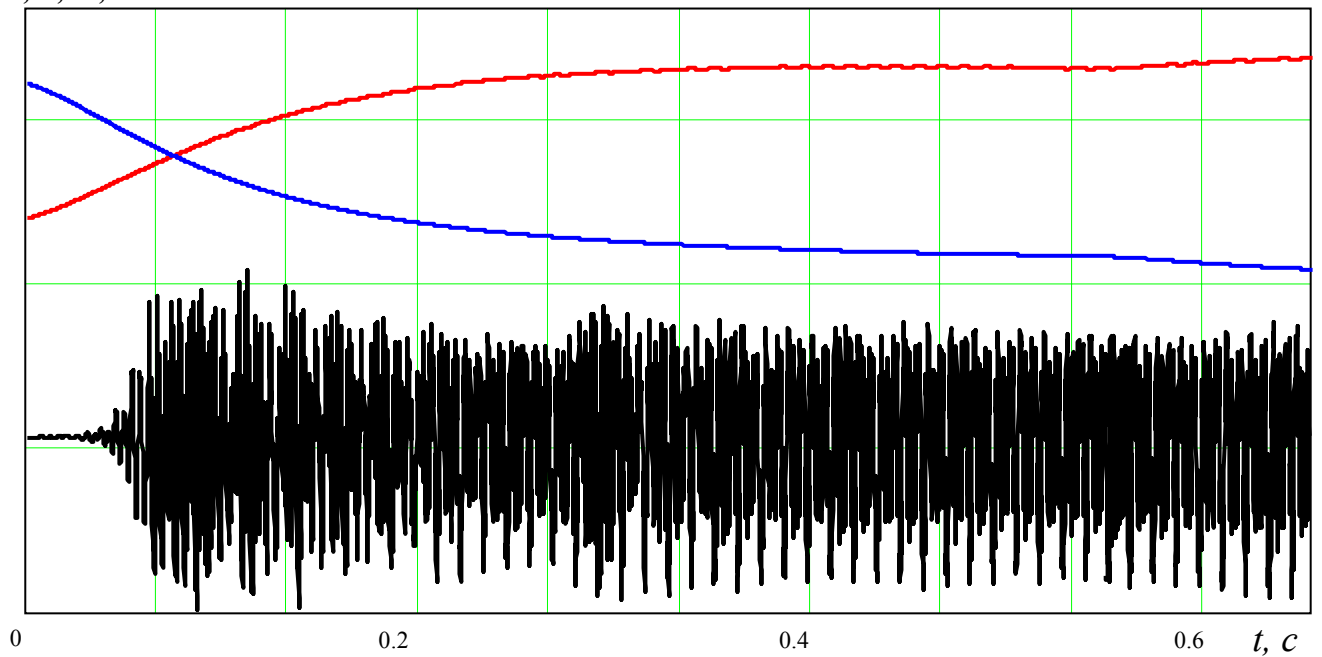
$U, C, B, \text{мФ}$ 

Рисунок 4.30 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «Э»

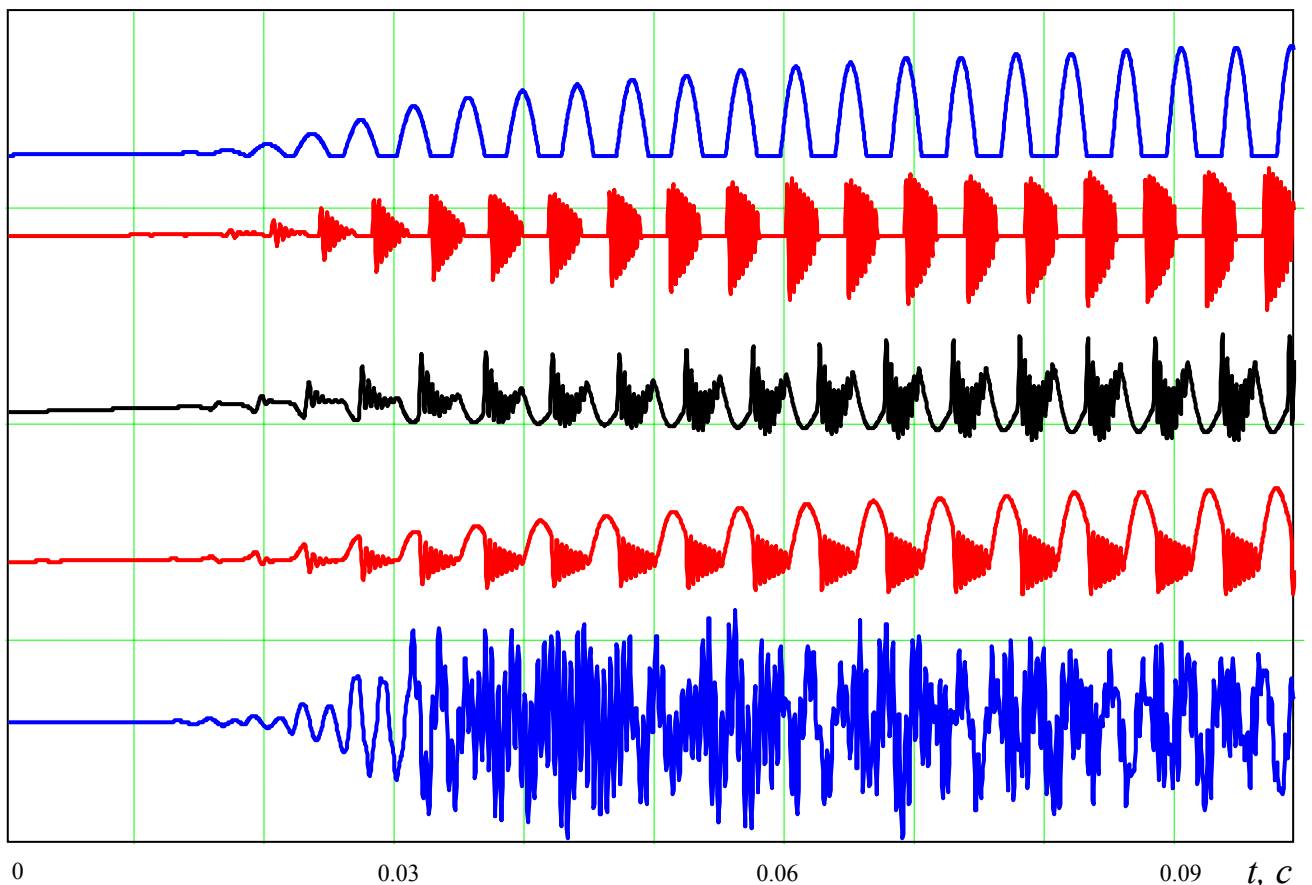
 $U, I, B, \text{мА}$ 

Рисунок 4.31 Формы сигналов на различных участках речеобразующего тракта для звука «Э». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (над-складочного и подскладочного), напряжение на выходе речеобразующего тракта

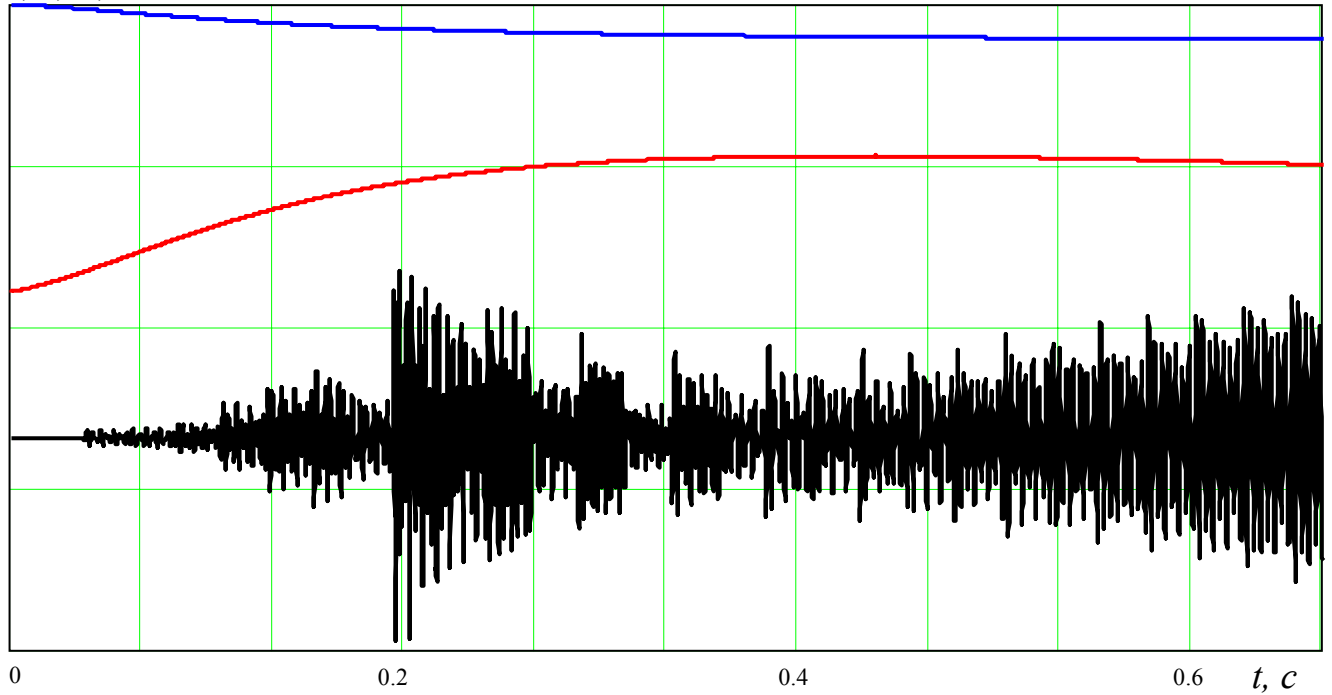
$U, C, B, \text{мФ}$ 

Рисунок 4.32 Избыточное давление в легких (---), емкость легких (----), давление на выходе речеобразующего тракта при произнесении звука «А» «пищеводный голос»

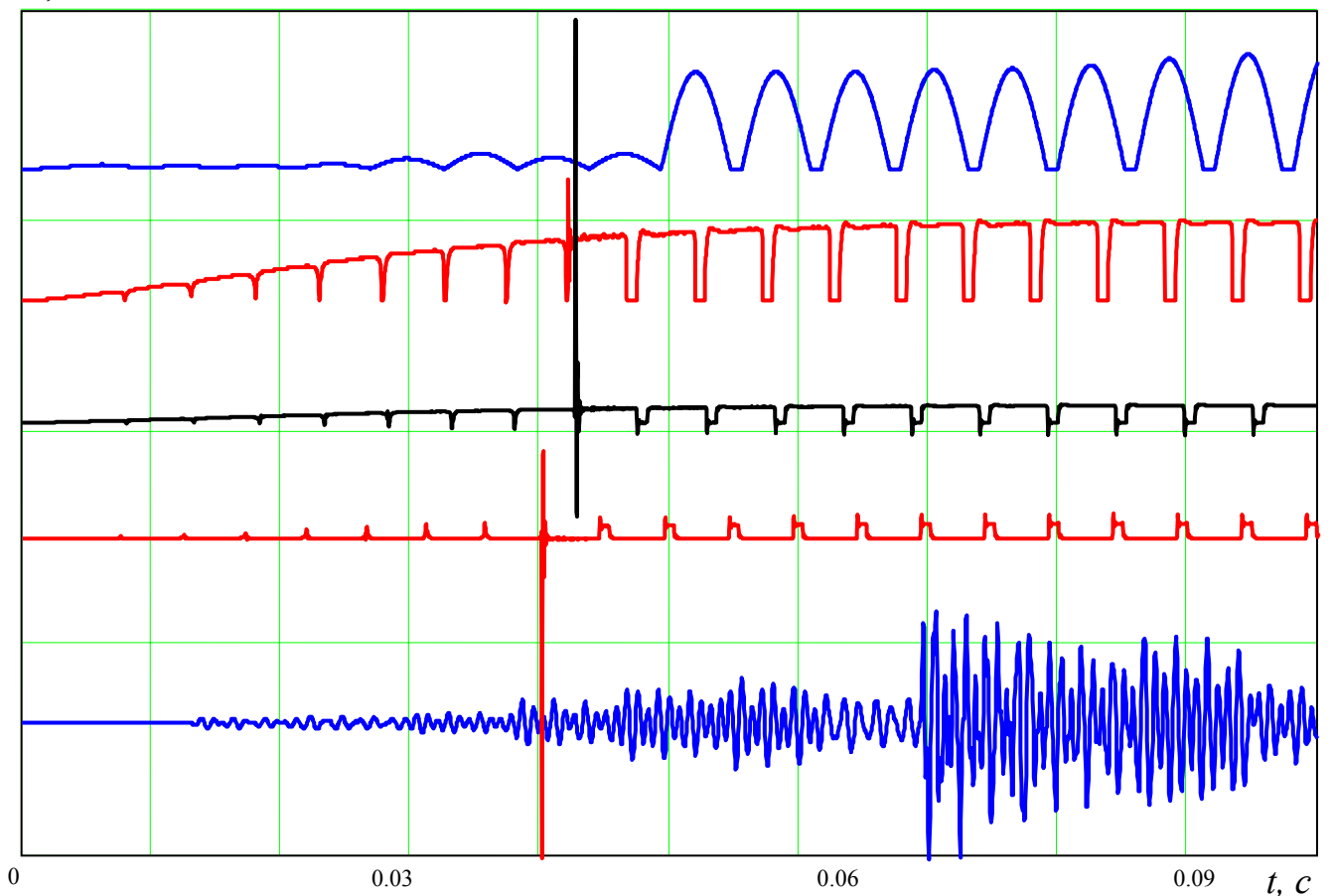
 $U, I, B, \text{мА}$ 

Рисунок 4.33 Формы сигналов на различных участках речеобразующего тракта для звука «И» последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение на глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта.

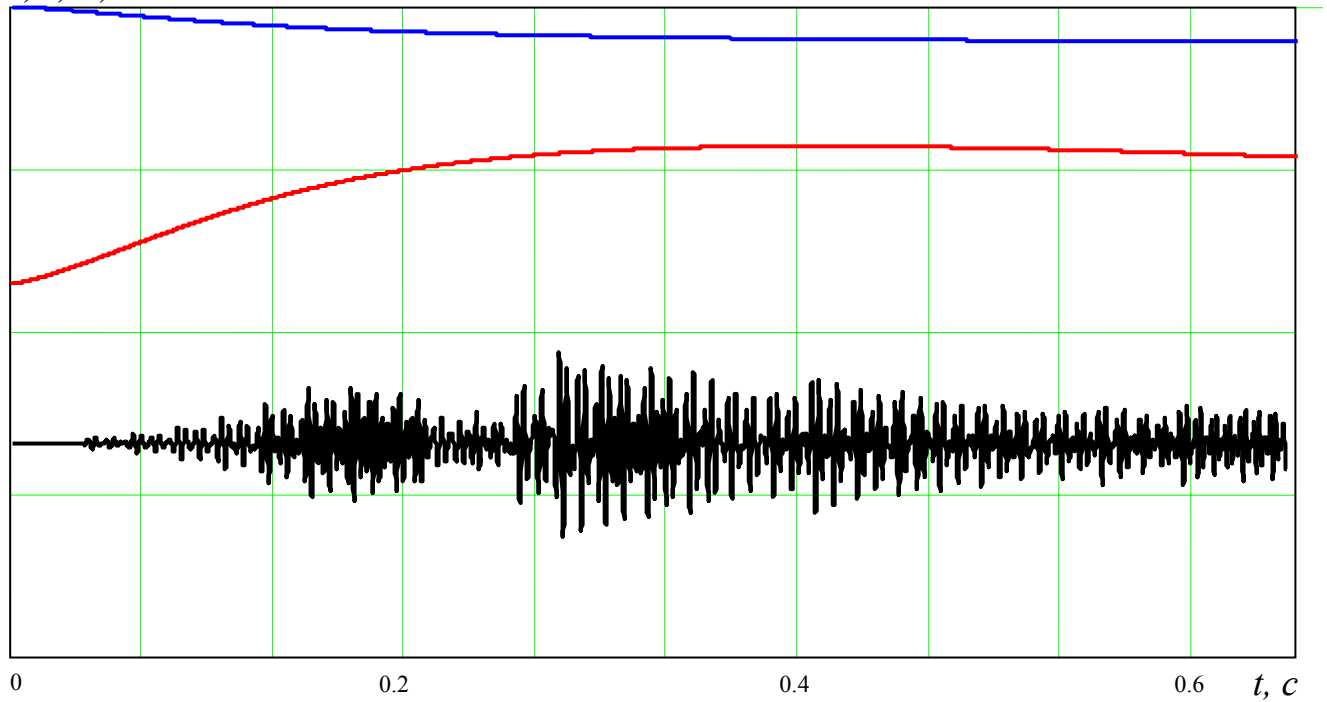
$U, C, B, \text{мФ}$


Рисунок 4.34 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «О» «пищеводный голос»
 $U, I, B, \text{мА}$

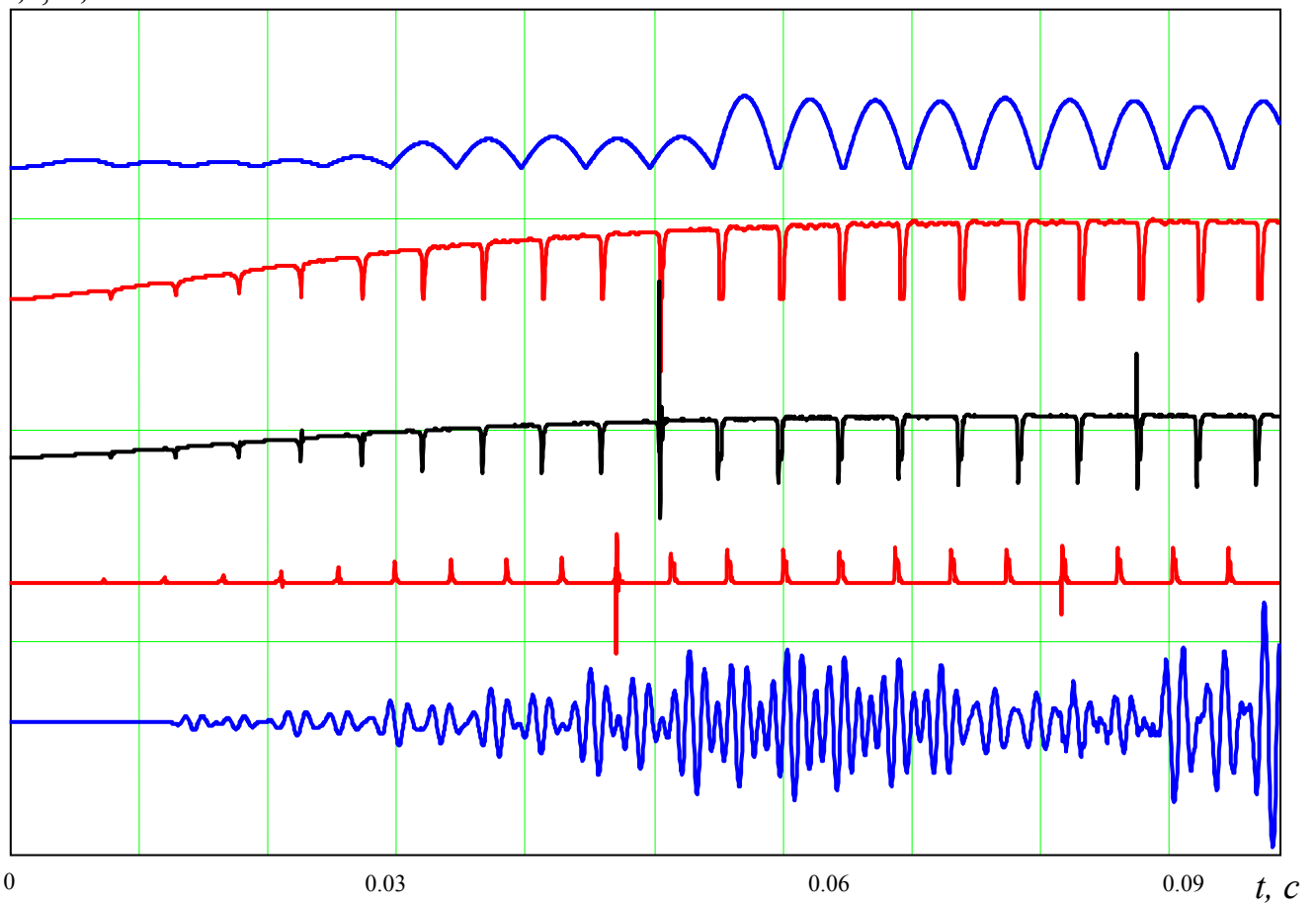


Рисунок 4.35 Формы сигналов на различных участках речеобразующего тракта для звука «О» «пищеводный голос». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

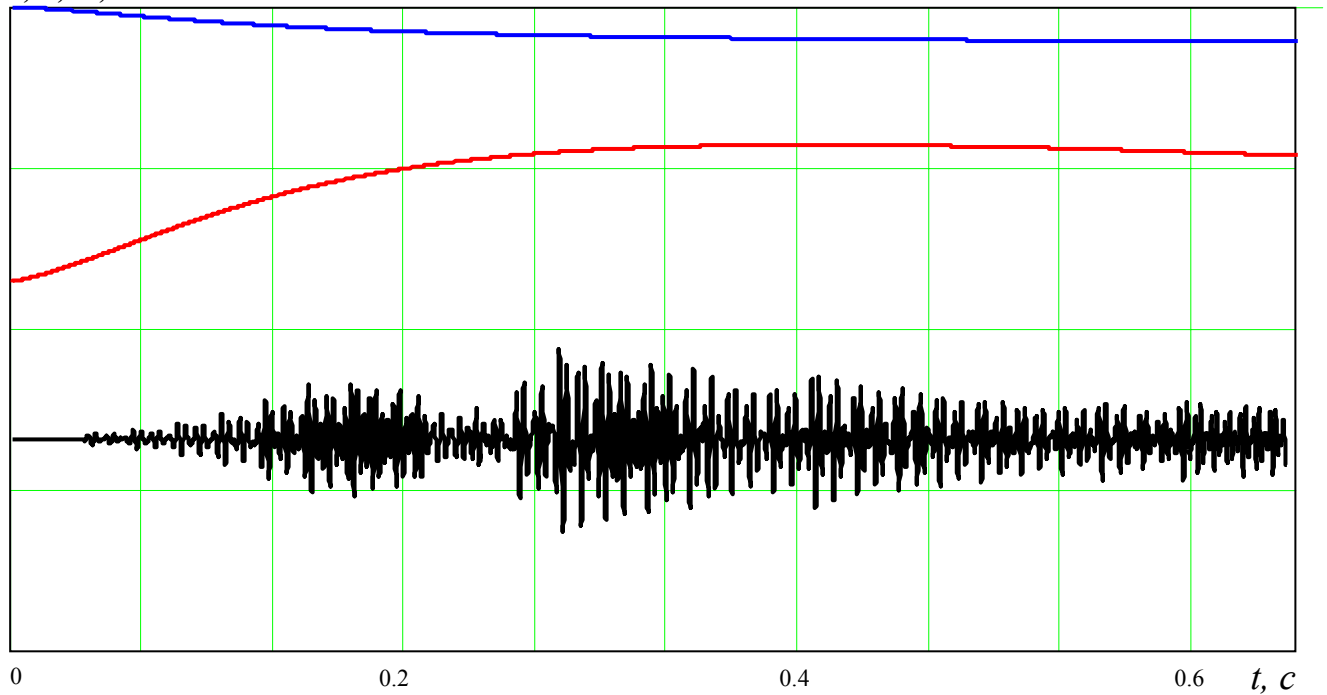
$U, C, B, \text{мФ}$


Рисунок 4.36 Избыточное давление в легких (---), емкость легких (---), давление на выходе речеобразующего тракта при произнесении звука «О» «пищеводный голос»
 $U, I, B, \text{мА}$

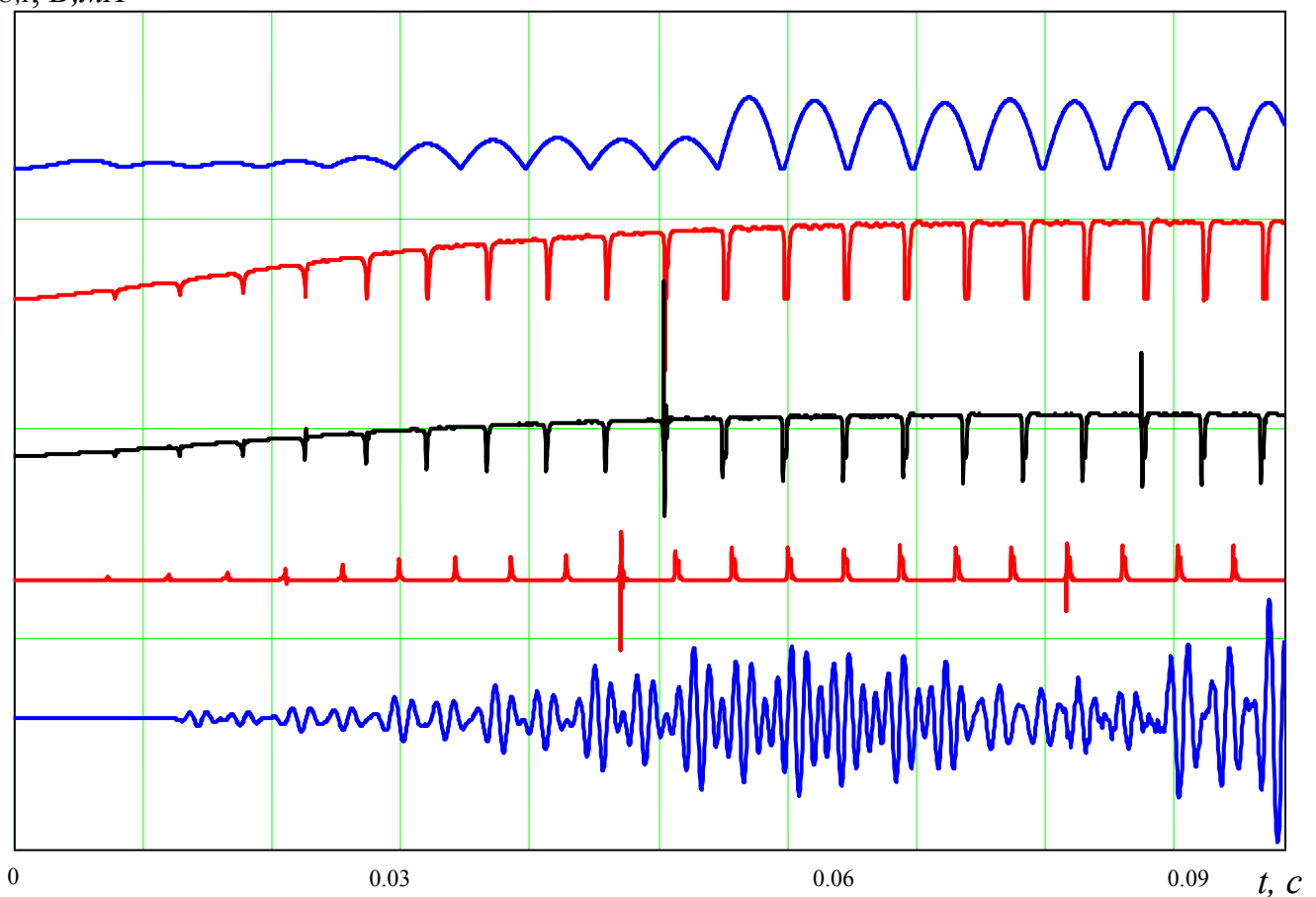


Рисунок 4.37 Формы сигналов на различных участках речеобразующего тракта для звука «О» «пищеводный голос». Последовательно: площадь раскрытия голосовых складок, ток через голосовые складки (I_s), напряжение в глоточном желудочке, разность давлений (надскладочного и подскладочного), напряжение на выходе речеобразующего тракта

Анализ временных диаграмм рис. 4.17 – 4.31 показывает, что форма давления в глоточном желудочке, а также ток через голосовые складки зависят от параметров легких и трахеи. Колебательные процессы в начале периода импульсов основного тона вызваны резонансными свойствами трахеи. Для данных экспериментов собственная частота колебаний в системе легкие - трахея составляет 2500 Гц. В более сложной модели трахеи число резонансов может быть значительно больше и давление на выходе голосовых складок иметь более сложную форму.

Анализ рис. 4.17 – 4.31 показывает, что форма давления на выходе голосовой щели, а также ток через раскрытые голосовые складки зависят от генерируемого звука и при определенных параметрах речеобразующей системы звук формируется уже в глоточном желудочке. Это наблюдается при малых значениях постоянных времени глоточного желудочка и звеньев речеобразующего тракта и малой индуктивности L_M .

При синтезе речевых сигналов нормальной речи емкость легких бралась в пределах 1200 – 2000 Ф, а для пищеводной на порядок меньше.

На данных примерах можно убедиться в том, что при фонации пищеводным голосом включаются компенсаторные механизмы голосообразующей системы.

В процессе обучения под руководством врача-логопеда необходимо сформировать колебательный орган, под которым должен быть воздушный пузырь как можно большего объема. Формирование воздушного пузыря на уровне V-VI шейных позвонков производится в несколько этапов, первый из них дыхательные упражнения. При помощи методики дыхательной гимнастики, которая описана в [38, 85, 90, 121], пациент формирует условный рефлекс заглатывания воздуха в пищевод. В связи с тем, что объем воздуха в воздушном пузыре существенно меньше, чем в легких, можно сравнивать речевые сигналы на выходе речеобразующего тракта и обратить внимание на

следующие различия для пищеводной речи по сравнению с нормальной при произнесении одних и тех же звуков:

1. Ток через раскрытые псевдоголосовые складки имеет прямоугольную форму. Это приводит к смещению энергии сигнала в высокочастотную область спектра.

2. Наблюдаются срывы колебаний практически на всех диаграммах (рис. 4.33, 4.35, 4.37).

3. Сигналы на выходе речеобразующего тракта имеют такую форму, что по временной картине можно, в ряде случаев, получить значение периода основного тона существенно ниже, чем периода колебаний голосовых складок или их эквивалента (рис. 4.19, 4.27, 4.33, 4.35, 4.37).

Одинаковый масштаб по времени позволяет сделать вывод, что каждый период работы голосовых складок можно фиктивно обнаружить несколько периодов основного тона, например, на рис. 4.19 на участке от 0,038 до 0,045 можно наблюдать один период работы голосовых складок и 2,5 квазипериодических участка, которые можно принять за периоды основного тона.

Результаты проведенного синтеза позволяют провести сравнение речевых сигналов, полученных на основе моделей образования нормальной и пищеводной речи, с реальными.

Однако необходимо отметить, что в данных моделях использованы упрощенные модели легких и трахеи. Поэтому, несмотря на высокую естественность сгенерированных речевых сигналов, следует ожидать, что спектральный состав их должен быть много беднее реальных. Но полученные результаты убедительно показывают, что взятое в основу модели положение о влиянии разности давлений в трахее и ротовой полости на движение голосовых складок является основным фактором, определяющим процесс фонации.

4.4 Исследование спектральных характеристик

Для вокализованных звуков речи огибающие частотного спектра являются одной из ключевых характеристик для их идентификации [108]. Поэтому у исследовательского комплекса (рис. 4.1) имеется возможность исследования интегральных характеристик, таких как спектральный состав речевого сигнала. Формантные частоты представляют резонансы речеобразующего тракта при произнесении звуков и отображаются в спектре речевого сигнала повышением интенсивности соответствующих компонентов.

Каждая фонема характеризуется собственным спектральным составом, что позволяет с некоторой точностью предсказать, какой звук был произнесен в данный момент времени. Однако необходимо учитывать, что идентифицировать как звук, так и диктора не всегда удастся по спектру речевого сигнала [7]. Это связано с тем, что в системе речеобразования формируется достаточно сложный сигнал, имеющий много частотных компонентов, интенсивность которых определяется текущими условиями возбуждения колебаний. Поэтому сравнительная оценка реальных и синтезированных сигналов имеет значение только в плане сопоставления их общей картины.

Спектральные свойства речевых сигналов анализировались при двух значениях добротности:

$$Q_i = \frac{Fp_i}{\Delta F_i},$$

где Fp_i – резонансная частота фильтра i – канала;

ΔFp_i – ширина амплитудно-частотной характеристики i -го – канала на уровне 0,7.

Распределение добротностей по каналам анализа определялось соотношением

$$Q_i = Q_{\min} \cdot e^{\alpha \cdot i}, \quad i = (0, I)$$

где α определялось из условия $Q_I = Q_{\max}$.

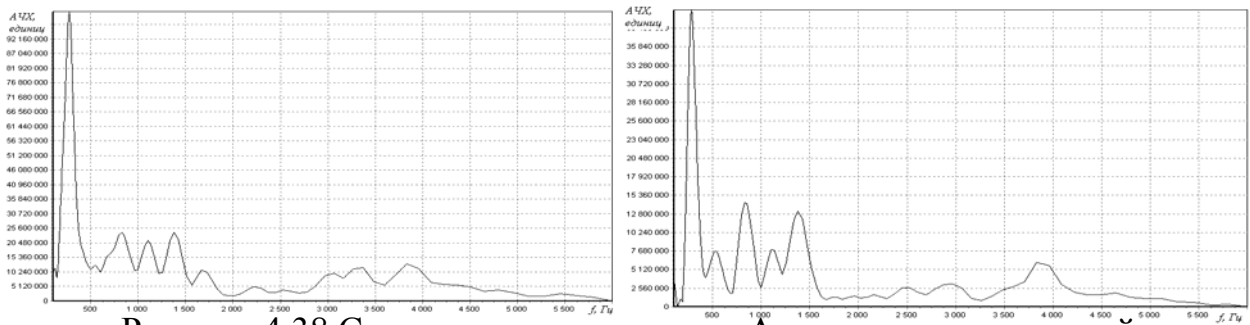


Рисунок 4.38 Спектр мощности звука «А» для разных значений добротностей: $Q=2$ и $Q=4$ (диктор - женщина)

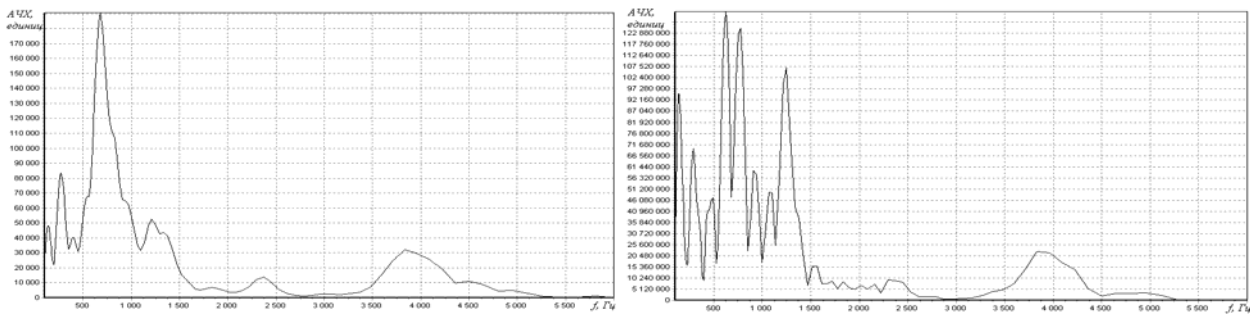


Рисунок 4.39 Спектр мощности звука «А» для разных значений добротностей: $Q=2$ и $Q=4$ (диктор - мужчина)

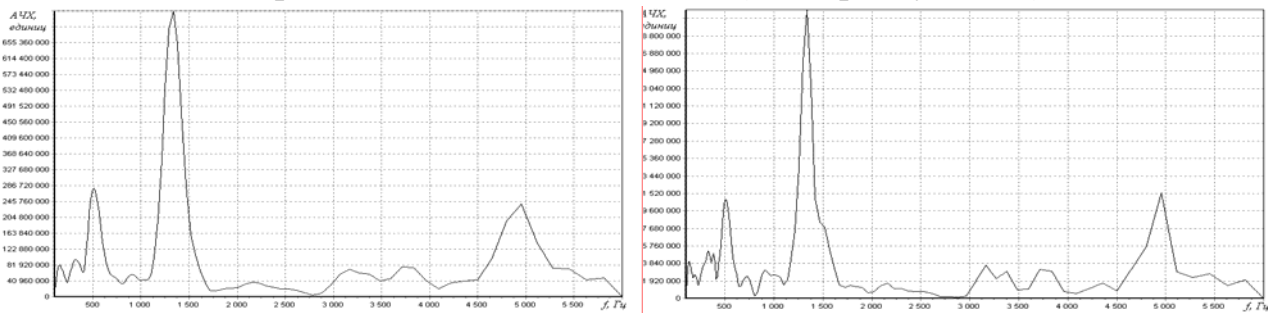


Рисунок 4.40 Спектр мощности звука «А» для разных значений добротностей: $Q=2$ и $Q=4$ (диктор - мужчина, фарингеальный голос)

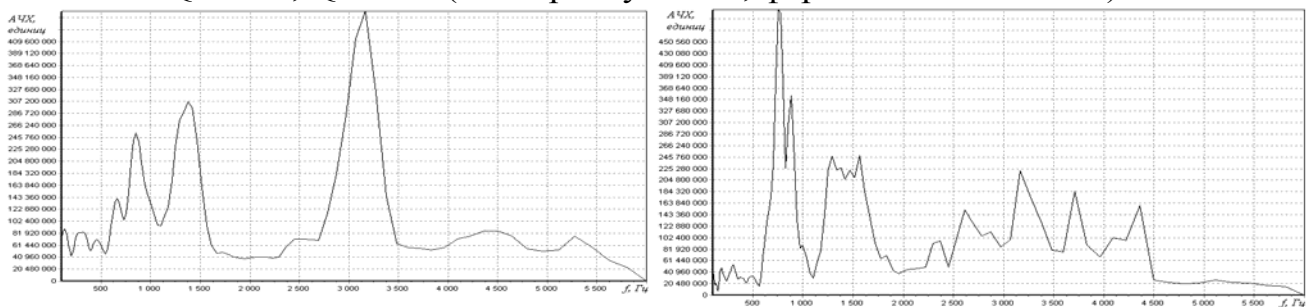


Рисунок 4.41 Спектр мощности звука «А» для разных значений добротностей: $Q=2$ и $Q=4$ (диктор - мужчина, пищеводный голос)

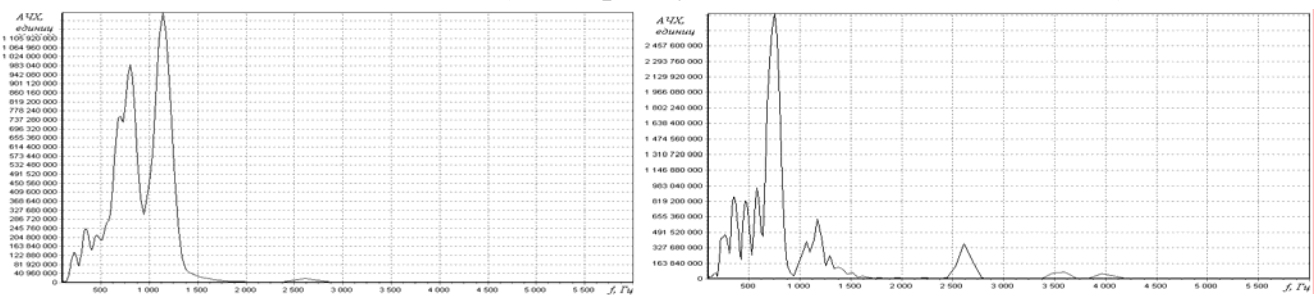


Рисунок 4.42 Спектр мощности звука «А» для разных значений добротностей: $Q=2$ и $Q=4$ (синтезированная речь, субъективно определяемая как мужской голос)

Спектры мощности рис. 4.38 – 4.42 были сняты для одинаковых моментов времени от начала звука для здоровых добровольцев.

Сравнение положения формантных частот для всех звуков практически совпадают и соответствуют расчетным значениям резонансов входного импеданса речеобразующего тракта. С учетом того, что синтезированные звуки субъективно воспринимаются как естественные, можно уверенно утверждать об адекватности моделей реальным речеобразующим системам.

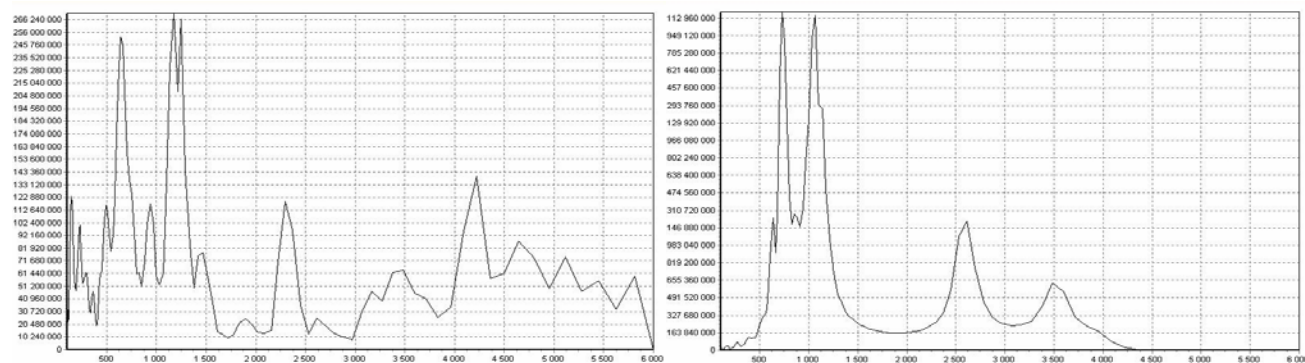


Рисунок 4.43 Спектр мощности звука «А» для пищеводной и синтезированной пищеводной речи при добротности $Q=4$ ч40

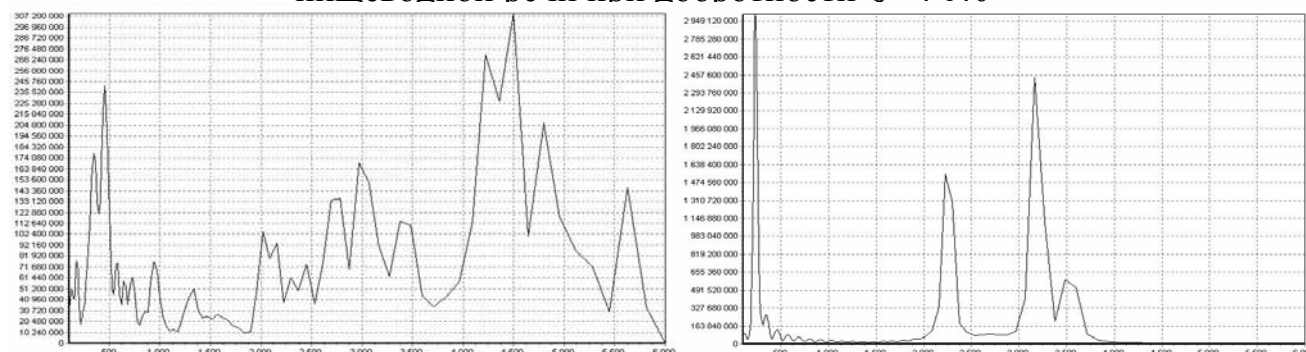


Рисунок 4.44 Спектр мощности звука «И» для пищеводной и синтезированной пищеводной речи при добротности $Q=4$ ч40

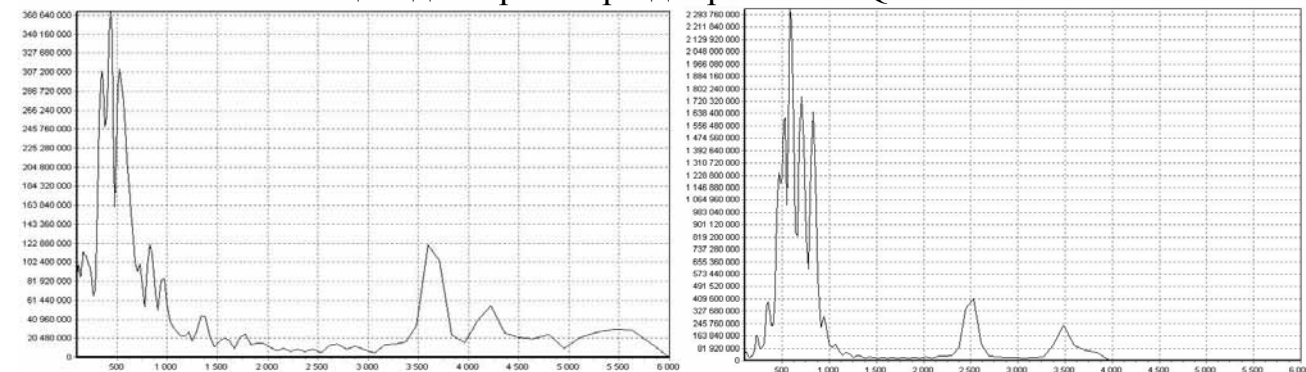


Рисунок 4.45 Спектр мощности звука «А» для пищеводной и синтезированной пищеводной речи при добротности $Q=4$ ч40

Вторая форманта имеет большую энергию для фрингеальной речи. Для пищеводной энергия сосредоточена в основном в районе четвертой и пятой формант. Эта закономерность характерна как для реальных, так и синтезированных сигналов. Необходимо отметить, что повышение разрешающей способности анализа по частоте ($Q=4\text{ч}40$) не позволяет разделить четвертую и пятую форманты. Однако практически для всех сигналов энергия сосредоточена в низкочастотной области. Это является характерной особенностью звучной речи.

Приведенные результаты анализа спектров мощности показывают, что они не дают возможности оценить особенности структуры речевого сигнала. Известно [7, 32, 36, 63, 83, 92, 96, 104, 105, 106], что распознавание по спектрам мощности речевых сигналов обладает низкой надежностью.

Наиболее качественно особенности структуры речевых сигналов выявляются с использованием механизмов маскировки динамических спектрограмм мощности. На рис. 4.46 - 4.51 приведены результаты применения последовательной и одновременной маскировки для различных речевых сигналов при $Q=1\text{ч}10$. На всех рисунках по оси абсцисс время, а ординат – частота.

На рис. 4.46 – 4.47 представлены результаты такой обработки речевого сигнала звука «А» для нормальной и пищеводной речи. Из этих рисунков следует, что для нормальной речи после последовательной маскировки четко выделяется периодическая структура, которая обусловлена частотой основного тона, т.е. периодичностью возбуждения. Эта периодичность после одновременной маскировки выделяется полосой в низкочастотной области. Легко обнаружить, что для пищеводной речи четкой периодичности нет. Это обусловлено высокой нестабильностью частоты основного тона. В высокочастотной области для нормальной и пищеводной речи четко выделяется после одновременной маскировки формантная структура сигнала.

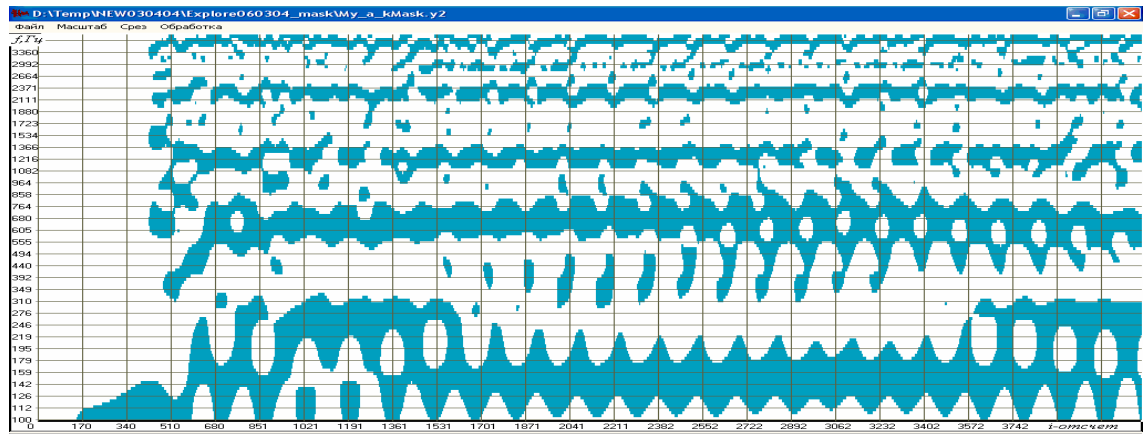


Рисунок 4.46 Звук «А» мужской голос, одновременная маскировка

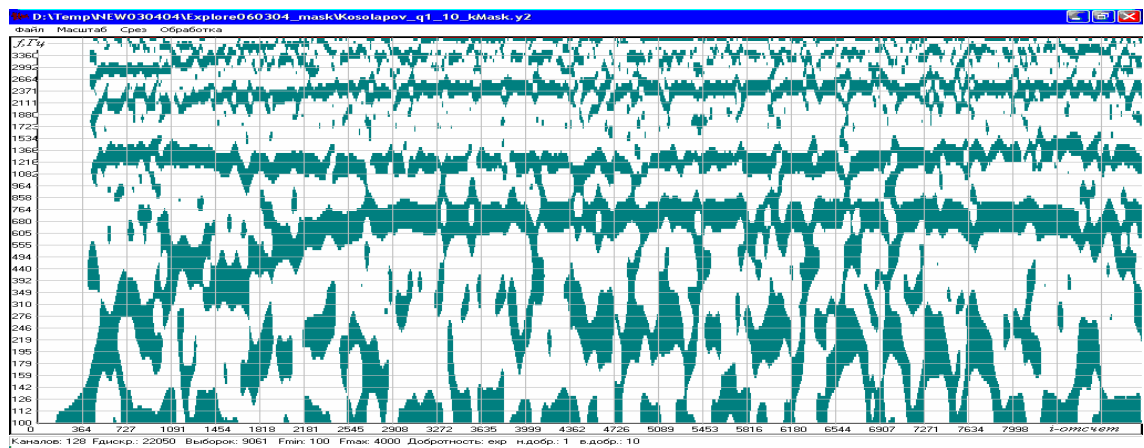


Рисунок 4.47 Звук «А» пищеводная речь, одновременная маскировка

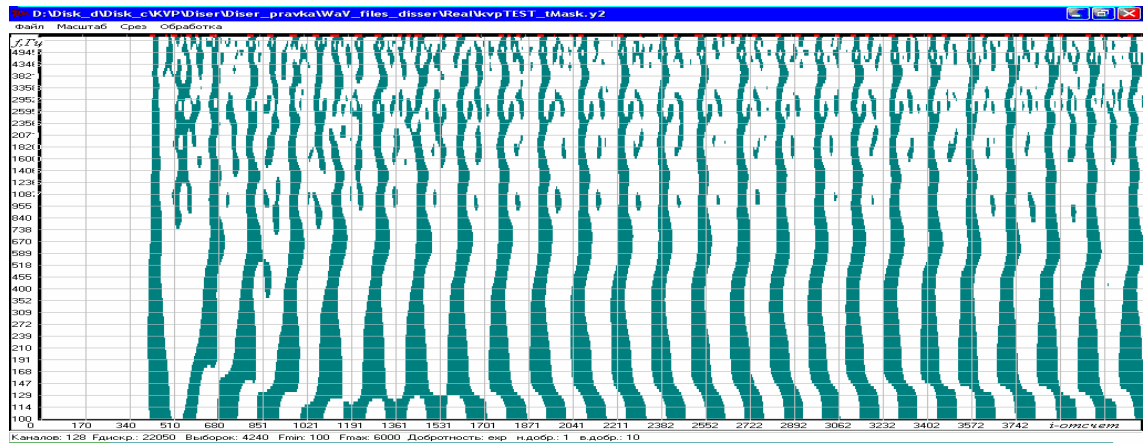


Рисунок 4.48 Звук «А» мужской голос, последовательная маскировка

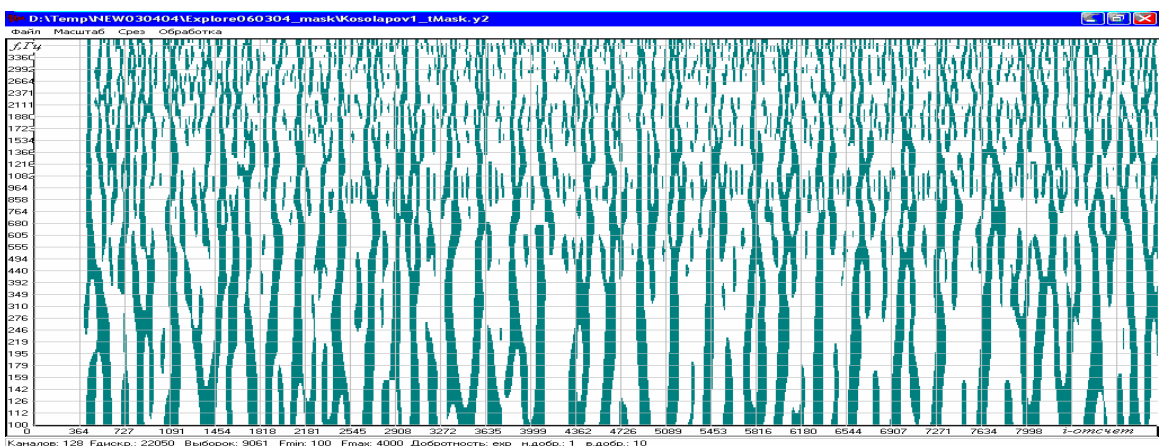


Рисунок 4.49 Звук «А» пищеводная речь, последовательная маскировка

На рис. 4.50 и 4.51 приведены результаты такой обработки для синтезированных сигналов нормальной речи звуков «А» и «И». По своей структуре полученные результаты совпадают с результатами, представленными на рис. 4.46 и 4.48. Это позволяет утверждать, что модель образования нормальной звучной речи адекватна реальной системе речеобразования.

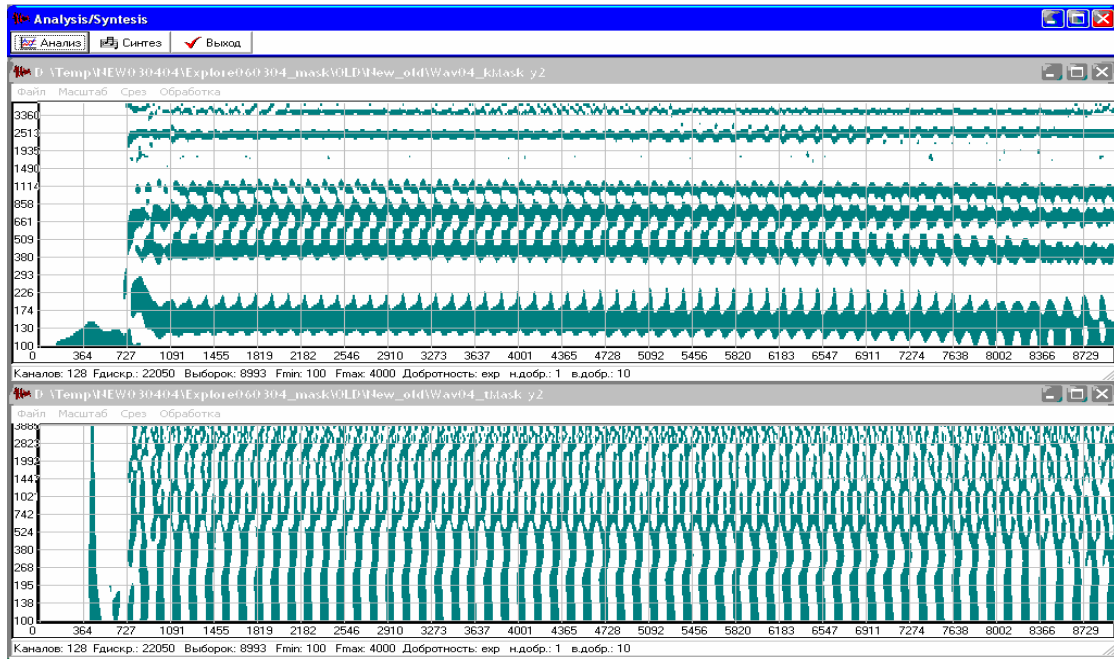


Рисунок 4.50 Звук «А» синтезированная речь, одновременная и последовательная маскировка

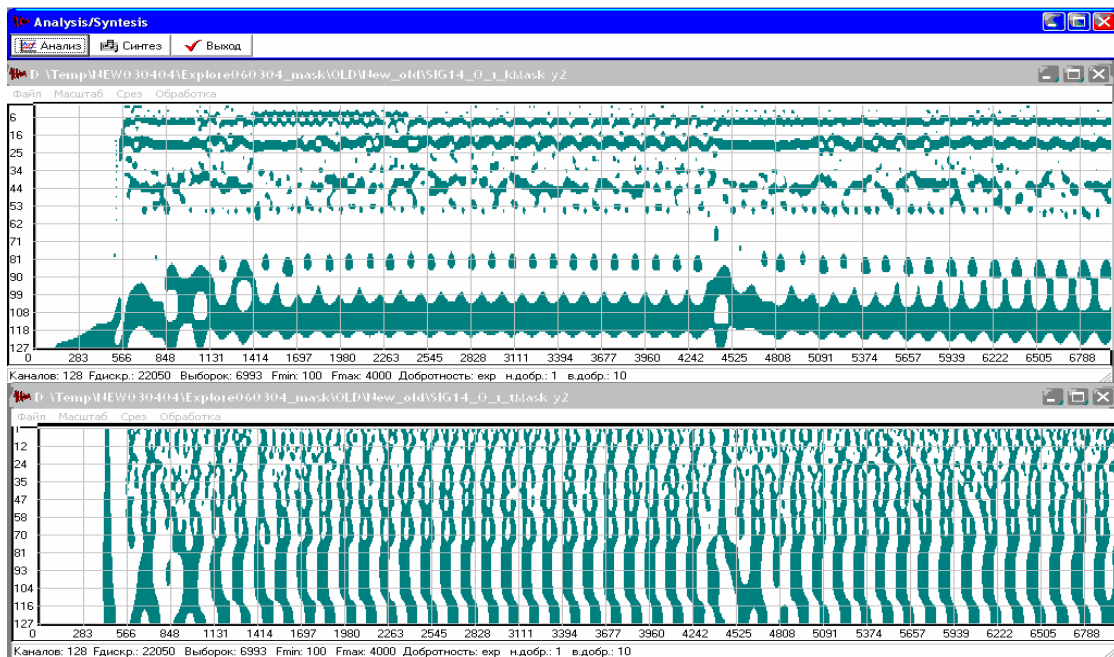


Рисунок 4.51 Звук «И» синтезированная речь, одновременная и последовательная маскировка

Особый интерес при постановки пищеводной речи представляет характер изменения частоты основного тона. Измерения частоты основного тона проводилось в одном низкочастотном канале, в котором имелся максимум энергии. Ясно, что в зависимости от величины Q этот максимум будет по-разному локализован, что демонстрируется на рис. 4.52 и 4.53, полученных для тех же сигналов, что и на рис. 4.46 и 4.47.

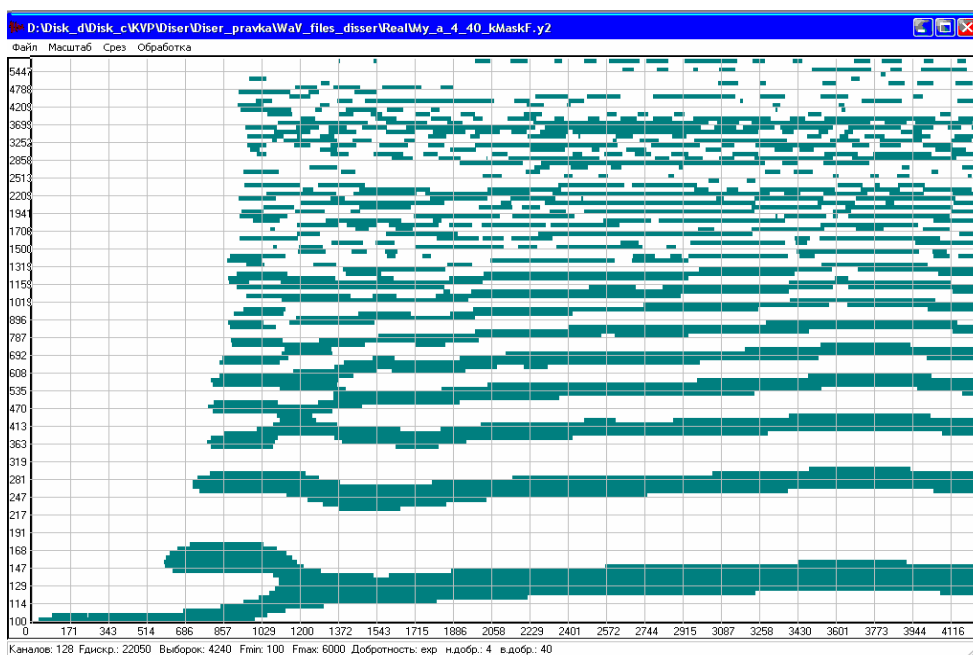


Рисунок 4.52 Звук «А» мужской голос, одновременная маскировка $Q=4$ ч40

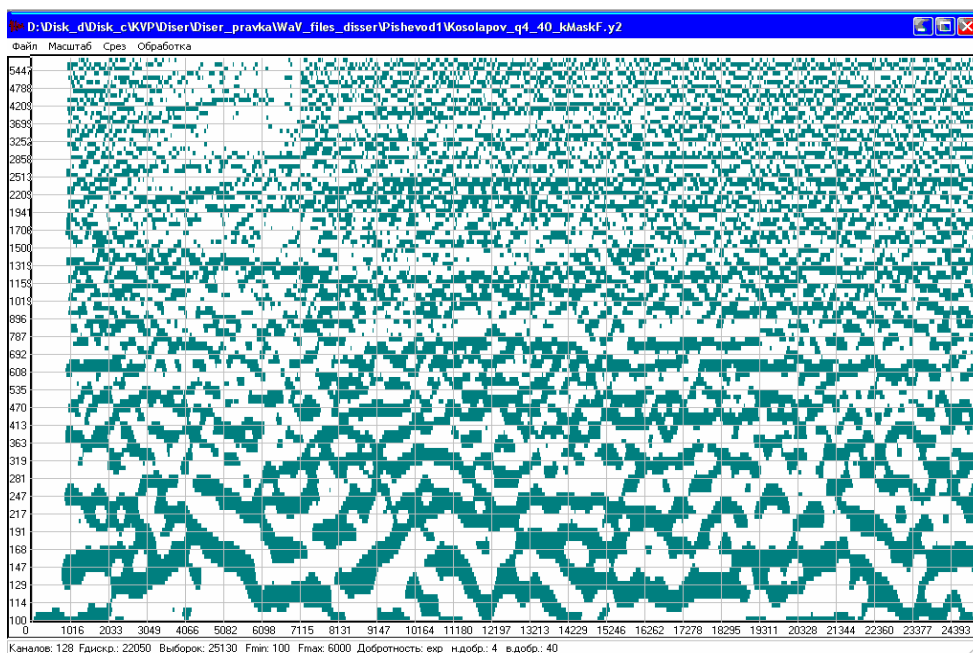
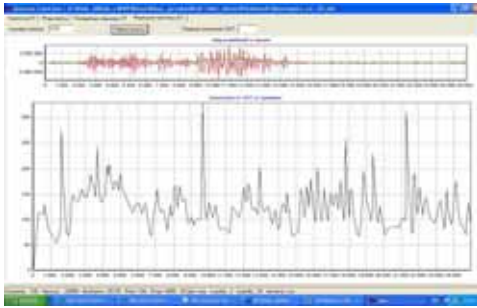
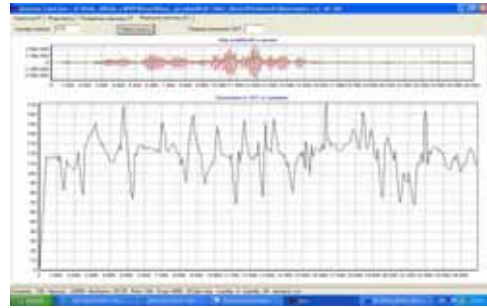


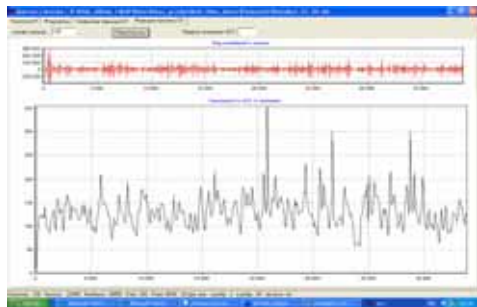
Рисунок 4.53 Звук «А» пищеводная речь, одновременная маскировка $Q=4$ ч40



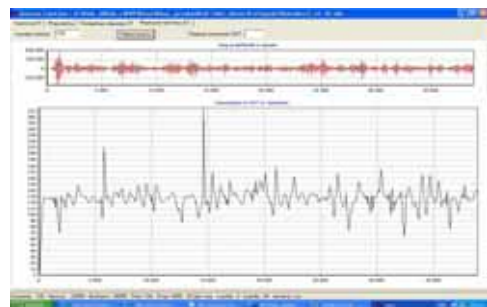
Пищеводная речь, звук «А» Q=2-20



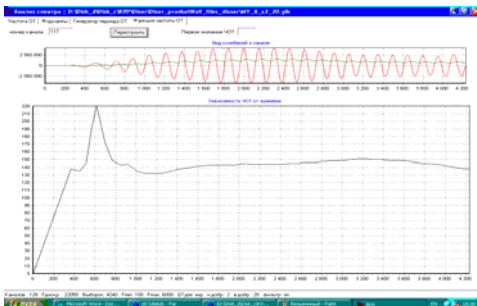
Пищеводная речь, звук «А» Q=4-40



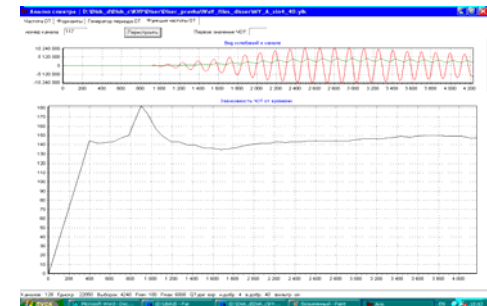
Фарингеальная речь, звук «А» Q=2-20



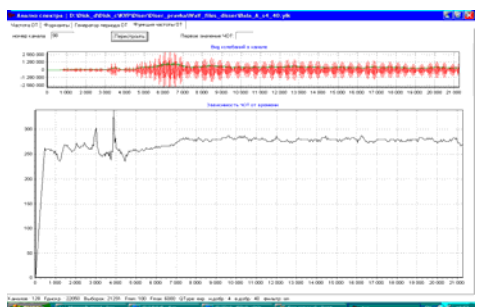
Фарингеальная речь, звук «А» Q=4-40



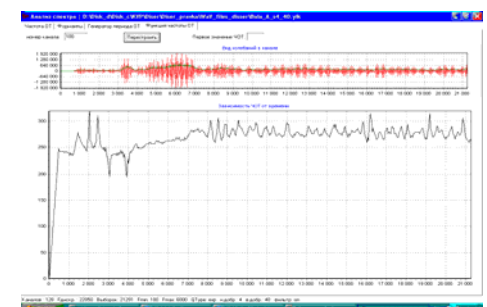
Мужской голос, звук «А» Q=2-20



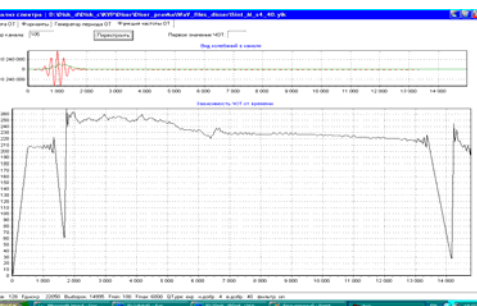
Мужской голос, звук «А» Q=4-40



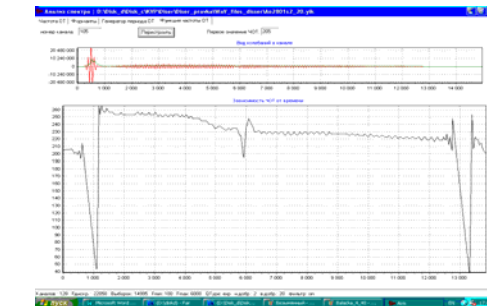
Женский голос, звук «А» Q=2-20



Женский голос, звук «А» Q=4-40



Синтезированный звук «А» Q=2-20



Синтезированный звук «А» Q=4-40

Рисунок 4.54 Частота основного тона

На рис. 4.54 представлены результаты измерения частоты основного тона для разных речевых сигналов, как реальных, так и синтезированных. Эти данные подтверждают результаты качественного анализа на основе динамических спектрограмм и на основе их обработки алгоритмами маскировки. Необходимо отметить, что как для реальных, так и синтезированных сигналов нормальной речи характерным является относительно высокая стабильность частоты основного тона во времени.

Для синтезированных сигналов пищеводной речи, так же как и для реальной, наблюдается нестабильность частоты основного тона. Однако эта нестабильность имеет несколько иной характер. По-видимому, это можно объяснить тем, что при образовании реальной пищеводной речи все-таки формируется обратная связь, которая поддерживает стабильность частоты основного тона. В рассматриваемой модели этой связи нет.

Таким образом, результаты проведенного сравнения показывают, что разработанные модели в плане образования звучной речи имеют много общего с реальными системами. Это позволяет целенаправленно проводить исследования как в плане оценки качества пищеводной речи, так и в плане совершенствования методики обучения пациентов пищеводной речи.

Выводы по главе:

Кратко описан комплекс для исследования структуры речевого сигнала. Данный комплекс был разработан аспирантом кафедры КИБЭВС ТУСУР Пономаревым А.А., но основные требования на его технические параметры были сформулированы на основе моделей речеобразования, рассмотренных в данной диссертации.

Проведено исследование структуры речевых сигналов, в частности его основной характеристики – частоты основного тона.

Измерения частоты основного тона на слогах реальных речевых сигналов косвенно подтверждают положение, положенное в основу моделей речеобразования, о том, что колебания голосовой щели определяются разностью давлений в трахее и ротовой полости.

Исследование структуры различных речевых сигналов, как реальных, так и синтезированных, позволяет утверждать, что предложенные модели адекватны, в первом приближении соответствуют реальным механизмам речеобразования.

Сравнение синтезированных и реальных речевых сигналов позволяет сделать вывод, что исследованные модели речеобразования адекватны реальным системам. Это подтверждается субъективным восприятием синтезированных сигналов, а также качеством их динамических спектрограмм.

Дальнейшее совершенствование моделей должно быть связано, по-видимому, с введением обратных связей по управлению собственной частоты колебаний голосовых складок, высотой гортани, давлением в легких в зависимости от генерируемого звука, его качества и т.п.

ЗАКЛЮЧЕНИЕ

В соответствии с целью диссертационной работы проведено:

Рассмотрение анатомии органов речеобразования при формировании нормальной и пищеводной речи. Составлены акустические схемы процессов фонации. Проанализированы существующие модели речеобразования. Предложен возможный механизм голосообразования, который явился основой постановки задачи.

Показано, что в качестве модели речеобразующего тракта может использоваться электрический аналог на сосредоточенных параметрах, что позволило построить модели голосообразования для нормальной и пищеводной речи, отличающиеся от известных тем, что они позволяют учитывать влияние перепада подскладочного и надскладочного давлений на процесс фонации. В результате чего на построенных моделях проведено исследование механизмов голосообразования с использованием двух гипотез: милоэластической и нейрорхностической, а также проанализировано влияние сил Бернулли. Выявлены качественные различия в процессах голосообразования нормальной и пищеводной звучной речи. Показано, что следует ожидать существенного повышения уровня высокочастотных компонентов в пищеводной речи, а также существенной неустойчивости частоты основного тона.

Проведена оценка параметров моделей образования звучной речи. Определены их размерности, и установлена связь между акустическими и механическими величинами и соответствующими величинами электрических аналогов. Выявлена особая роль гортанного желудочка. Установлены допустимые значения для постоянных времени трахеи и звеньев речеобразующего тракта, при которых обеспечивается близкое к реальному раскрытие голосовой щели. Определены граничные условия для коэффициентов, учитывающих влияние сил Бернулли. Исследованы области, в которых возникают устойчивые колебания в моделях образования нормальной

и пищеводной речи, что позволило синтезировать речевые сигналы произвольной длительности.

Проведен анализ сигналов пищеводной и нормальной речи. Проведено их сравнение с реальными речевыми сигналами. Результаты сравнения показывают, что модели отражают основные процессы при речеобразовании. Результаты синтеза убедительно доказывают, что главное значение при возникновении колебаний голосовых складок имеет разность давлений. Для стабильности колебаний голосовых складок существенное значение имеет как емкость, в которой формируется избыточное давление, так и емкость гортанного желудочка.

Полученные результаты легли в основу комплексного исследования пищеводной речи в отделении опухолей головы и шеи НИИ онкологии Томского научного центра СО РАМН и позволили улучшить качество речи пациентов, а также сократить время реабилитационного периода по формированию органов в I физиологическом сужении пищевода.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аксютин И.В., Карневская Е.Б., Лобанов Б.М., Марченков М.А. Многофакторная модель ритмики и ее реализация при синтезе речи по тексту. // АРСО 15, Таллин, 1989 – 347с., с.135-136
2. Балацкая Л.Н., Кицманюк З.Д., Чойзонов Е.Л., и др. Реабилитация больных опухолями головы и шеи – Томск: НТЛ, 2003. – 296с.
3. Бегун П. И., Шукейло Ю. А. Биомеханика: Учебник для вузов. - СПб: Политехника, 2000. - 463 с.
4. Бондаренко В.П., Безходарнов И.В., Коцубинский В.П. Анализ и синтез речи в системах человеко-машинного взаимодействия // 2-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-2-96)»: Тезисы докладов. – Новосибирск: 1996 - 146с. с.116-117
5. Бондаренко В.П., Коцубинский В.П. Синтез речи по печатному тексту // 3-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-3-97)» Тезисы докладов, Красноярск, 1997- 232 с, с156-157
6. Бондаренко В.П., Коцубинский В.П., Маркивский И.О. Модель синтеза речевых сигналов // Автоматизация проектирования, идентификация и управление в сложных системах. Сб.статей/ Под ред. В.П.Тарасенко. - Томск: НТЛ, 1997. –250 с., с.88-102
7. Бондаренко В.П., Коцубинский В.П., Маркивский И.О. Исследование структуры речевых сигналов гласных звуков // Методы и алгоритмы автоматизации технологически процессов. Сб.статей/ Под ред. В.П.Тарасенко. - Томск: Изд-во Том. Ун-та, 1995.-252с. с.72-81
8. Бондаренко В.П., Маркивский И.О. Оценка параметров системы фильтров модели анализа речевых сигналов / Методы и алгоритмы автоматизации технологических процессов. - Томск.: ТГУ, 1995. – 252с., с.61-70

9. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Управление просодией при синтезе речи // Управляющие и вычислительные системы. Новые технологии: Материалы межвузовской научно-технической конференции: Вологда: ВоГТУ, 2000. -210с., с.211-213
10. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Влияние легких на колебания голосовых связок при генерации вокализованных звуков // Сборник трудов XI сессии российского акустического общества. - М.: ГЕОС 2001. Том 3. 214с. 56-59с.
11. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Иерархическая система синтеза речи по правилам/ Сборник: Автоматическое и автоматизированное управление сложными системами: Сб. статей / Под. Ред. В.П.Тарасенко. - Томск: Изд-во Том. ун-та, 1998. - 236с, с.218-227
12. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Иерархические структуры распознавания и синтеза речи // Интеллектуальные автоматизированные системы проектирования, управления и обучения. Сб.статей/ Под ред. В.П.Тарасенко. - Томск: ТГУ, 2000. 266с. 115-125с.
13. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Модель управления просодией при синтезе речи // Теория и практика речевых исследований (АРСО-99). Материалы конференции. - Москва.: МГУ им. М.В. Ломоносова 1999 - 160с. с.37-38
14. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Нелинейные механизмы в работе голосового источника // Системные проблемы качества, математическое моделирование информационных, электронных и лазерных технологий. // Материалы международной конференции и Российской научной школы: ч.: 2. – М.: Радио и связь, 2002 - 144с. с. 93-95
15. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Нелинейные механизмы в образовании звучной речи // Труды Нижегородской акустической научной сессии. / Под. ред. С.Н. Гурбатов. - Нижний Новгород: Часть 3. – ТАЛАМ, 2002, 409с., с 200-204

16. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Распознавание и синтез речи в структуре диалоговых систем // 2-ая всероссийская конференция «Теория и практика речевых исследований». Материалы конференции. – М.: МГУ им. М.В. Ломоносова 2001 160с. с.142-150
17. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Синтез речевого сигнала по печатному тексту // Автоматическое и автоматизированное управление сложными системами. Сб.статей/ Под ред. В.П.Тарасенко. – Томск: Изд-во ТГУ, 1998. 238с. 204-218с.
18. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В. Формирование просодии при синтезе речевых сигналов // Интеллектуальные автоматизированные системы проектирования, управления и обучения. Сб.статей/ Под ред. В.П.Тарасенко. – Томск: ТГУ, 2000 - 266с. 126-130с.
19. Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В., Пономарев А.А., Климович М.К. Проблемы понимания речи в задачах анализа и синтеза // Интеллектуальные системы в управлении, конструировании и образовании. Выпуск 2 / Под ред. А.А. Шелупанова. – Томск: STT, 2002. – 232с. с. 44-52
20. Бондаренко В.П., Пономарев А.А. Исследование структуры речевого сигнала // Интеллектуальные системы в управлении, конструировании и образовании / Под ред. А.А.Шелупанова. – Томск: STT, 2001. - 224с., с. 17-20
21. Бондарко Л.В. Звуковой строй современного русского языка. - М.: Просвещение, 1977 – 175с.
22. Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики. - СПб.: Издательство СПб университета, 1991 – с.212
23. Бондарко Л.В., Вербицкая Л.А., Щербакова Л.П. Об определении места ударения в слове. // Известия АН СССР. Серия литературы и языка, 1973 - т.32, вып.2. с.141-153.
24. Бракерст Э., Миноу М., Течнер У. Система с трехуровневой программой для преобразования текста в речь. // Электроника – 1983. - № 8, с. 36-42.
25. Быков Ю.С. Теория разборчивости в линиях связи. – М.: Оборониздат, 1954, –196с.

26. Быков Ю.С. Теория разборчивости речи и повышение эффективности радиотелефонной связи. – М.-Л.: Госэнергоиздат, 1959 – 351с.
27. Вемян Г. В. Качество телефонной передачи и его оценка – М.: Связь, 1970. - 224 с.
28. Винницкий А.С. Модулированные фильтры и следящий прием ЧМ сигналов. – М.: Советское радио, 1969. – 548 с.
29. Ворсано Д. Кодирование речи в цифровой телефонии // Сети и системы связи №1 – 1996 – с 12-15
30. Галунов В.И., Станкевич С.А. О модели речеобразования // VII всесоюзная акустическая конференция по физической и технической акустики 1-6 февраля 1971г.: Сб. докл. / Под ред. Л.Л. Мясникова – Ленинград: 1973. - 340с. с. 42-44.
31. Гельднер К., Кубик С. Нелинейные системы управления: Пер с нем. - М.: Мир, 1987. – 386с.
32. Гринорян А.А., Цимель Г.И. К опознаванию гласных звуков и неоднородных гласных по переходным сегментам гласных. // Проблемы передачи информации 1969 т.5 №4 с.38-49.
33. Годин Р. Система речевого ввода и вывода информации обзор // Электроника, 1983 - № 8 - с.27-29.
34. Гудонавичус А. П., Кемешис П.П., Читавичус А.Б. Распознавание речевых сигналов по их структурным свойствам. – Л.: Энергия 1977-- 148с.
35. Денисов А.А., Нагорный В.С. Пневматические и гидравлические устройства автоматики – М.: Высшая школа, 1978 - 216с.
36. Динамические спектры речевых сигналов. / Под ред. Дергач М.Ф. – Львов: Высшая школа, 1983 – 294с.
37. Дмитриев Л. Б., Телелева Л. М., Таптапова С. Л., Ермакова И. И. Фониатрия и фонопедия – М.: Медицина, 1990 - 134с.
38. Ермолаев В.Г., Лебедева Н.Ф., Морозов В.П. Руководство по фониатрии. – М.: Медицина, 1970. – 268 с.
39. Жинкин Н.И. Механизмы речи. – М.: АПН РСФСР, 1957. – 372 с.

40. Журавлев А.П., Павлюк Н.А. Язык и компьютер. М.: Просвещение, 1989
41. Зевеке Г.В., Ионкин П.А., Страхов С.В. Основы теории цепей. – М.: Энергоатомиздат, 1989 – 528с.
42. Златоустова Л.В. Интонация и просодия в организации текста //Звучащий текст. – М.: Институт научной информации по общественным наукам, 1983 – с. 11-21.
43. Златоустова Л.В. Фонетическая структура слова в потоке речи - Казань: Издательство Казанского университета, 1962 – 214с.
44. Златоустова Л.В. Фонетические единицы русской речи. - М.: Издательство МГУ, 1981 – 206 с.
45. Златоустова Л.В., Кодзасов С.В., Кривнова О.Ф., Фролова И.Г. Алгоритмы преобразования орфографических текстов в фонетическую запись. - М.: МГУ, 1970 – 312с.
46. Златоустова Л.В., Кузнецов В.Б., Отт А. Длительность гласного как результат взаимодействия 2-х факторов: ритмической структуры слова и его положения в синтагме // АРСО-13, Новосибирск: НГУ, 1984 - с. 3-5.
47. Златоустова Л.В., Потапова Р.К., Трунин-Донской В.Н. Общая и прикладная фонетика – М.: Издательство МГУ, 1986. – 304с.
48. Исакович М.А. Общая акустика. – М.: Наука, 1973 – 495с.
49. Кейтер Дж. Компьютеры – синтезаторы речи. – М.: Мир, 1985 - 237с.
50. Коперда Ф. Система распознавания речи. /Компьютеры: Справочное руководство т.3.- М., 1986 - гл. 30.- с.358-391.
51. Коренбаум В.И., Кулаков Ю.В., Тагильцев А.А. Акустические эффекты в системе дыхания человека при форсированном вдохе // Акустический журнал, 1997, том 43, № 1 - с. 78-86.
52. Коротков Ф.А., Богачева А.В., Добрынин А.Н., д.р. Струйная техника в автоматике – М.: Энергия, 1977 - 168с.
53. Косарев Ю.А. Естественная форма диалога с ЭВМ. – М.: Машиностроение, Ленингр. Отд., 1989. – 142 с.
54. Коцубинский В.П. Определение характеристик акустического фильтра для

- формирования гласных звуков // 6-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-6-99)» Тезисы докладов Омск 2000 - 282с. 75-76с.
55. Коцубинский В.П. Особенности машинного и программного обеспечения обработки речи // 4-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-4-98)» Тезисы докладов Барнаул 1998 – 230с. 131-132с.
 56. Коцубинский В.П. Особенности формирования звучной речи // Интеллектуальные системы в управлении, конструировании и образовании / Под ред. А.А.Шелупанова. – Томск: STT, 2001. – 224с., с. 33-44
 57. Коцубинский В.П. Программы синтеза и распознавания английской речи // Сибирская школа молодого ученого: Труды научно-практической конференции студентов, аспирантов и молодых ученых, 21-23 декабря 1998, г. Томск с. 56 – с 14-15.
 58. Коцубинский В.П., Мещеряков Р.В. Модель системы речеобразования // Измерения, автоматизация и моделирование в промышленности и научных исследованиях: Межвузовский сборник //Под. Ред. Г.В. Леонова. – Бийск: Алт.гос.тех.ун-та, 2001.-276с. с.48-52.
 59. Коцубинский В.П., Мещеряков Р.В. Особенности формирования звучной речи в норме и патологии // 7-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-7-2001)»: Тезисы докладов Барнаул 2001. Ч.2 300с. с.189-191
 60. Коцубинский В.П., Мещеряков Р.В. Речевая диалоговая система // 7-ая международная научно-практическая конференция «Природные и интеллектуальные ресурсы Сибири (Сибресурс-7-2001)»: Тезисы докладов Барнаул 2001. Ч.2 300с. с.191-192
 61. Коцубинский В.П., Удалов Д.В. Форматный синтезатор речевого сигнала // Интеллектуальные автоматизированные системы проектирования, управления и обучения. Сб.статей/ Под ред. В.П.Тарасенко. - Томск: Изд-во ТГУ, 2000. 266с. 131-136с.

62. Кривнова О.Ф. Количественная оценка воздействия супraseгментных факторов на длительность ударных гласных в синтагме. // АРСО-13, 1984, НГУ, с. 6-7.
63. Кривнова О.Ф., Гладков С.М., Чардин И.С., Бабкин А.В. Узнайте его по голосу // Компьютерра. - 1997 - №49 - с.40-43
64. Крил П. API для распознавания речи для WINDOWS 95 И WINDOWS NT //Computerworld №20 1996, - с48-97
65. Кузнецов В.Б., Отт А.В. Автоматический синтез речи. Алгоритм преобразования "буква-звук" и управление длительностью речевых сегментов - Таллин: "Валгус" 1989 – 135с.
66. Кутуков Г.П., Матюнин С.Н. Состояние исследований в академических и учебных институтах страны в области автоматического распознавания и синтеза речи // Информационные технологии. - 1998. - №6. - с.21-26
67. Кучеров В.Я., Лобанов Б.М. Синтезаторы речи в системах массового обслуживания. – М.: Радио и связь, 1983 – 132с.
68. Кюннап Э.Ю. Синтез речи. – Таллин: Академия наук Эстонской ССР, 1988
69. Кюннап Э.Ю. Синтезаторы речевых сигналов. - Таллин: Валгус, 1975 – 98с.
70. Ла-Салль Ж., Левшец С. Исследование устойчивости прямым методом Ляпунова: Пер. с англ. – М.: Мир, 1964. –168 с.
71. Левшец С. Устойчивость нелинейных систем автоматического управления: Пер. с англ. – М.: Мир, 1967 – 184 с.
72. Лепендин Л.Ф. Акустика: Учебное пособие для вузов. – М.: Высшая школа, 1978. – 448с.
73. Лобанов Б.М. и др. Одноплатный модуль синтеза речи по печатному тексту. // АРСО-14. Каунас, 1986. – с.54-55.
74. Лобанов Б.М. Микроволновой синтез речи по тексту // Сб. научных трудов «Анализ и синтез речи» / Под ред. Лобанова Б.М., - Минск: АН БССР.Ин-т техн.кибернетики; 1991 86с., с. 57-73.
75. Лобанов Б.М. Оптимальный модуль синтеза речи // АРСО-12. Киев, 1982. – с.516-521.

76. Лобанов Б.М. Принципы автоматического синтеза интонационных структур. //АРСО-10, Тбилиси, 1978 – 159с., с.36-42
77. Лобанов Б.М. Теоретические основы систем речевого общения. // Искусственный интеллект: в 3 кн. Кн.1 Системы общения и экспертные системы: Справочник / под ред. Э.В. Попова М.: Радио и связь, 1990. - 464с., ил.
78. Лойцянский Л.Г. Механика жидкостей и газов. – М.: Наука, 1978 – 736с.
79. Майлс С. Подводная медицина. – М.: Медицина, 1971 – 328с.
80. Математическая лингвистика. Сборник переводов. /Под ред. Шрейдера Ю.А., Ревзина И.И., Лахути Д.Г. и Финна В.К., - М.: Мир, 1964
81. Мессарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем /Пер. с англ. под ред. И.Ф.Шахнова - М.: Мир, 1973 – 311с.
82. Методы автоматического распознавания речи: В 2-х кн.Кн.1//Авт.: У.А. Ли, Э.П.Нейбург, Т.Б.Мартин и др./; Под ред.У.Ли; Пер.с англ. О.В.Александровой под ред. А.А.Воронова -М.:Мир,1983. – 327 с
83. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи. /Под редакцией М.А. Сапожкова/- М.: Радио и связь. 1987 - 168с.
84. Морозов В.П. Искусство резонансного пения. Основы резонансной теории и техники – М.: МГК, ИП РАН., 2002 – 426с.
85. Морозов В.П. Разборчивость вокальной речи как функции высоты основного тона голоса //Акустический журнал. 1964 т. 10 вып 3 с. 376-380.
86. Мясников Л.Л. Автоматическое распознавание звуковых образов – Л.: Энергия, 1970 – 183 с.
87. Мясникова Е.Н. Объективное распознавание звуков речи – Л.: Энергия, 1967, – 148 с.
88. Ольшанский В.О., Решетов И.В., Новожилова Е.Н., Дворниченко В.В. Восстановление голоса у больных раком гортани после ларингэктомии при помощи голосовых протезов // Сибирский онкологический журнал – 2003. № 2, - с. 40-44.

89. Основы автоматического управления. Под редакцией В.С.Пугачева. - М.: Наука, 1967 – 680с
90. Основы логопедии. / Филичева Т.Б., Чевелева. Н.А., Чиркина Г.В. - М.: Просвещение, 1989. - 223с.
91. Основы пневмоавтоматики / В. Н. Дмитриев, В. Г. Градецкий.-М.: Машиностроение, 1973. – 360 с.
92. Пайнен Д. Распознавание речи: завтра уже наступило? /Мир ПК, 1995 - №13 - с.38-44.
93. Петленко Б.И., Бутырский Л.С. Речевая связь в искусственных атмосферах. – М.: Связь, 1978 – 216с.
94. Пиотровский Р.Г. и др. Математическая лингвистика. М.: Высшая школа, 1977 – 387с.
95. Пономарев А.А., Рогодзинская Е.А., Бондаренко В.П. Модель одновременной маскировки // Интеллектуальные системы в управлении, конструировании и образовании Выпуск 3/ Под ред. А.А.Шелупанова. – Томск: СТУ, 2004. - 216с, с. 167-174
96. Покровский Н.Б. Расчет и измерение разборчивости речи. М.: Связь издательство, 1976 – 391с.
97. Потапова Р.К. Речевое управление роботом. - М.: Радио и связь, 1989. - 246с., ил.
98. Потапова Р.К. Речь: коммуникация, информация, кибернетика: Учеб. пособие для вузов. - М.: Радио и связь. 1997 – 528с.
99. Потапова Р.К. Система делимитативных средств звучащего текста //Звучащий текст. М.: Институт научной информации по общественным наукам, с.22-60.
100. Потапова Р.К. Тайны современного Кентавра – М.: Радио и связь, 1992 – 246с.
101. Правила и алгоритм преобразования орфографической записи на русском языке в фонетическую транскрипцию. А.В. Кельманов, Н.В. Соломатина, А.Г. Хайретдинова и др //Прикладные системы искусственного интеллекта

(вычислительные системы) - 1995 - Вып. 153

102. Пресс Р.И., Плоткин Е.О. Импульсные устройства струйной техники – Минск, Наука и техника, 1977 – 208 с.
103. Проблемы прочности в биомеханики. / Под ред. Образцова И.Ф. – М.: Высшая школа, 1988 – 311с.
104. Рабинер Р. Теория и применение цифровой обработки сигналов: Пер. с англ. - М.: Мир, 1978. – 848 с.
105. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь. 1981 – 376 с.
106. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. М.: Радио и связь, 1981 -224с.
107. Рамишвили Г.С., Чикоидзе Г.Б. Криминалистическое исследование фонограмм речи и идентификация личности говорящего. - Тбилиси: Мецниереба, 1991, – 265с.
108. Сапожков М.А. Речевой сигнал в кибернетике и связи. - М.: Связьиздат, 1963 – 452с
109. Светозарова Н.Д. Интонационная система русского языка Л.: Издательство ЛГУ, 1982 - 175с.
110. Светозарова Н.Д., Щербакова Л.П. Роль изменения частоты основного тока в восприятии ударения в изолированных словах и предложениях./ АРСО - V1, Таллин: Институт кибернетики, 1972
111. Сдвижков О.А. MathCAD-2000: Введение в компьютерную математику. – М.: Дашков и К⁰, 2002. – 204 с.
112. Складов О.П. Нелинейная модель акустического голосового источника // Акустический журнал, 1997, том 43, № 5, с. 675-680.
113. Сорокин В.Н. Временные параметры элементов русской речи. Речевая информатика - М.: Наука, 1974 – 223с.
114. Сорокин В.Н. Некоторые вопросы построения общей модели речеобразования. Труды АРСО-8, Львов, 1974 – с.97-100.

115. Сорокин В.Н. О роли подглоточной области в процессе речеобразования. В кн. Проблемы построения систем понимания речи. – М.: Наука, 1980, с. 125-136
116. Сорокин В.Н. Синтез речи - М.: Наука, 1992 -392с.
117. Сорокин В.Н. Теория речеобразования. – М.: Радио и связь. 1985- 312с.
118. Сорокин В.Н., Трифоненко В.П. Об автокорреляционном анализе речевого сигнала // Акустический журнал 1996, том 42 № 3 с.418-425
119. Структура и семантика текста. – Воронеж: Изд. Воронеж. Ун-та, 1988
120. Тавернье К. Схемы синтезаторов речи: Заставьте ваши устройства говорить! – М.: ДМК Пресс, 2001 – 176 с.
121. Таптапова С.Л. Восстановление звучной речи у больных после резекции или удаления гортани. – М.: Медицина, 1985 – 96с.
122. Теория и практика речевых исследований (АРСО-99). Материалы конференции – М.: МГУ им. М.В. Ломоносова 1999 –160с.
123. Трунин-Донской В.Н. Автоматический синтез звучащего текста //Звучащий текст. – М.: Институт научной информации по общественным наукам 1983 – с. 218-250с.
124. Тугоухость. Под ред. Н.А. Преображенского. - М.: Медицина 1978 – 440 с.
125. Фант Г. Акустическая теория речеобразования . Пер. с англ. – М.: Наука, 1964. – 264 с.
126. Фант Г. Анализ и синтез речи: Пер. с англ. //Под ред. Загоруйка Н.Г. – Новосибирск: Наука, 1970 – 312с.
127. Фланаган Дж. Л. Анализ, синтез и восприятие речи: Пер. с англ. – М.: Связь, 1968 –396 с.
128. Флейшман Б.С. Элементы теории потенциальной эффективности сложных систем. - М.: Советское радио, 1971 – 223с.
129. Фонетика спонтанной речи // Под ред. Н.Д. Светозаровой. - Л.: ЛГУ, 1988.
130. Фониатрия и фонопедия // Дмитриев Л.Б., Теляева Л.М., Таптапова С.Л., Ермакова И.И. - М.: Медицина, 1990 – 272с.

131. Фортран: Программирование учебной пособие. / Под. Ред. Чл. –кор. АН УССР Е. Л. Ющенко. – 2-е изд., перераб. и доп. – Киев: Вища школа. 1980. – 400 с.
132. Хоэнштейн Л. Вывод звуковой информации: воспроизведение музыки и речи. / Компьютеры: Справочное руководство т.3. - М., 1986 - гл.29-с.342-357.
133. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации: Пер. с нем. – М.: Связь, 1971 – 256с.
134. Юссон Р. Певческий голос. – М.: Медицина 1974 - 262с.
135. Badin P., Bailly G., Raybadi M., Segebarth C. A 3-dimensional linear articulatory model based on MRI data // Proc. 3th ESCA/COCOSDA Int Workshop on Speech Synthesis, 1998, 249-254
136. Badin P., Engwall O. An MRI study of Swedish fricatives: coarticulatory effects // CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling, 5th Seminar on Speech Production, Kloster Seeon, Germany, 2000
137. Baer T., Gore J.C., Gracco L.C., Nye P.W. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels // J. Acoust. Soc. Am. 90 (2), 1991, 799-828.
138. Bagshaw P.C. An investigation of acoustic events related to sentential stress and pitch accent, in English - Speech Communication 1993 13(3-4): 333-342
139. Bagshaw P.C., Hiller S.M., Jack M.A. Enhanced pitch tracking and the processing of F0 contours for computer aided intonation teaching. Pages 1003-1006 of: Proc. 3rd. European Conference on Speech Communication and Technology, vol. 2. Berlin 1993
140. Berg Jw., van den Folia phoniatr., 1957, 9, 2.
141. Bondarenko V.P., Kotsubinsky V.P., Metharykov R.V. Physiological processes lungs model at speech formation // Processing 'SPECOM'2000'. International workshop Speech & Computer. -St. Petersburg. 2000. 102с. с79-82

142. Bondarenko V.P., Kotsubinsky V.P., Metharykov R.V. Speech signal synthesis model // Processing 'SPECOM'2001'. International workshop Speech & Computer. –Moscow.: MSLU 2001. 262p. pp.251-254
143. Bondarenko V.P., Kotsubinsky V.P., Metharykov R.V. Speech synthesis in dialog systems // Processing 'SPECOM'2002'. International workshop Speech & Computer. -St. Petersburg. 2002. 124p. pp.103-105
144. Bondarenko V.P., Kotsubinsky V.P., Metharykov R.V., Ponomarev A.A. The speech signal analysis and problems of speech understanding // Processing 'SPECOM'2002'. International workshop Speech & Computer. -St. Petersburg. 2002. 124c. c73-74
145. Bondarenko V.P., Moor V.R., Chabanets A.N. The analysis of speech perception mechanisms on the models of auditory system, Proceedings XIth ICPhS The eleventh international congress of phonetic sciences (volume 2), 1987.
146. Harper P., Kraman S.S., Pasterkamp H., Wodicka G.R. An acoustic model of the respiratory tract // IEEE Transactions on biomedical engineering, vol. 48, NO. 5, May 2001.
147. Husson R. La voix chantée. Paris, - 1960.
148. Husson R. Physiologie de la phonation. Paris, - 1962
149. Klatt D. H., Review of Text-To-Speech Conversion for English // Journal of the Acoustic Society of America (JASA), vol. 82 N3, 1987, pp. 737-793
150. Klatt, D.H. Interaction between two factors that influence vowel duration // JASA, 54, 1973. pp. 1102-1104
151. Knipper A. A SYLLABLE APPROACH TO THE SPEECH INFORMATICS. V.1 Se 2.5., p.56-59
152. Kohler Klaus J. Microprosody in segment perception Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V.1, Se 4.1, p.80-83
153. Kosarev Yu., Osipov A., Machovikov A. Natural spoken language understanding: integration of heterogeneous knowledge. Processing "SPECOM'98". International workshop SPEECH AND COMPUTER. - St. Petersburg. - 1998.

154. Kosarev Yu.A. The model of oral speech semantic interpretation: quantitative processing and integration of acoustic, syntactic, semantic and pragmatic data. Proc. German Acoustics Conference DAGA-94. Drezden, 1994, p 1281-1284
155. Kowtko J.C. The function of Intonation in Spontaneous and Read Dialogue. Proceedings of the Fifth European Conference on Speech Communication and Technology (Eurospeech 97), September 1997, Rhodes, Greece
156. Neil P. McAnglus Todd. Guy J.Brown A computation model of prosody perception. Proceeding of the International Conference on Spoken Language Processing (ICLSP-94), Yokohama, Japan, 18th-22nd September 1994, pp. 127-130
157. Nikolayeva T. The typology of sentence intonation systems. Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V.6, Se 102.3, p.106-109
158. Ottesen Georg E. Adding natural prosody to a phoneme synthesizer. Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V. 3, Se 52.4, p. 260-261
159. Ruch W.J. An Atlas of Common Subluxations of the Human Spine and Pelvis. CRC Press 1996 178p
160. Sanders E., Taylor P. Using statistical models to predict phrase boundaries for speech synthesis. Proceedings of the Fourth European Conference on Speech Communication and Technology (Eurospeech 95), September 1995, Madrid, Spain
161. Sproat R., Taylor P., Tanenblatt M., Isard A. A Markup Language for Text-To-Speech Synthesis. Proceedings of the Fifth European Conference on Speech Communication and Technology (Eurospeech 97), September 1997, Rhodes, Greece.
162. Story B., Titze I., Hoffman E. Vocal tract area functions from magnetic resonance imaging // J. Acoust. Soc. Am. 100 (1), 1996.
163. Svetozarova N.D. LINGUISTIC FACTORS IN SENTENCE STRESS (EVIDENCE FROM RUSSIA) Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V.6 Se 102.4, p 110-113

164. Taylor P. Synthesizing Intonation using the RFC Model. In Proc. ESCA Workshop on Prosody, Lund, Sweden, 1993
165. Taylor P. The Rise/Fall/Connection model of intonation. Speech Communications, forthcoming, 1994, 15: pp 169-186
166. Taylor P., Isard A. SSML: A Speech Synthesis Markup Language. Speech Communication 21
167. Taylor Paul, Shimodaira Hiroshi, Isard Stephen, King Simon, Kowtko Jaqueline Using Prosodic Information to Constrain Language Models for Spoken Dialogue.
168. Taylor, P.A. The Tilt Intonation Model, Proceedings of the Fifth International Conference on Spoken Language Processing (ICSLP 98), November 1998, Sydney, Australia.
169. Titze I., Story B., Hoffman E. Vocal tract area functions for an adult female speaker based on volumetric imaging // J. Acoust. Soc. Am. 104 (1), 1998, 471-487.
170. Williams B. Diphone synthesis for Welsh Proceedings of the IOA, vol. 16, part 5 (1994): 359-365.
171. Wright H., Taylor P. Modeling Intonational Structure Using Hidden Markov Models ESCA workshop on intonation, September 1997, Athens, Greece.
172. Zlatoustova L., Kozlenko N., Khitina M., Zakharov L., Automatic word stress detector. Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V.1, Se 4.5, p. 96-99
173. Zlatoustova L.V., Kedrova G.Y. Perspective and acoustic characteristics of emotions: A typological Research Based on the Material of Languages with Different Structures. Proceeding Xith ICPhS, 1987 Tallinn, Estonia, V.3, Po 2.6, p.218-221

ПРИЛОЖЕНИЕ 1

Поперечные сечения речеобразующего тракта от голосовых складок до рта (в см^2), при произнесении соответствующих звуков, приведенных из книги Сорокина В.Н. Синтез речи [116].

$x, \text{см}$	А	Э	И	О	У	Ы	С	Х	Ф
0,00	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75	2.75
0,44	2.17	2.17	2.17	2.17	2.23	2.27	1.53	2.17	2.17
0,88	1.47	1.47	1.47	1.47	1.56	1.67	0.88	1.47	1.47
1,32	1.03	1.03	1.03	1.03	1.16	1.27	0.68	1.03	1.03
1,76	0.81	0.81	0.81	0.81	0.89	0.94	3.38	0.81	0.81
2,20	0.65	0.65	0.65	0.65	0.69	0.79	4.06	0.65	0.65
2,64	0.72	0.72	0.72	0.72	0.67	0.65	3.32	0.72	0.72
3,08	1.14	1.16	1.16	1.15	0.74	0.69	3.23	1.14	1.17
3,52	4.38	4.86	4.86	4.52	1.37	0.83	3.09	4.23	5.13
3,96	4.11	4.65	4.65	4.27	5.9	2.63	3.7	3.98	4.95
4,40	3.73	4.06	4.06	3.82	5.84	5.31	3.81	3.64	4.38
4,84	3.25	3.59	3.59	3.35	5.39	5.21	3.72	3.11	3.78
5,28	2.27	3.27	3.27	2.52	4.88	4.72	3.66	2.54	3.46
5,72	1.08	3.13	3.13	1.61	3.93	4.35	3.67	2.55	3.31
6,16	1.87	3.95	3.95	2.21	3.5	3.7	3.77	3.52	4.2
6,60	1.31	3.87	3.87	1.66	3.87	3.58	3.99	3.46	4.03
7,04	0.87	3.81	3.81	1.18	3.42	4.41	4.34	3.35	3.95
7,48	0.61	3.8	3.8	0.74	2.87	4.68	4.64	3.21	3.91
7,92	0.86	3.84	3.84	0.56	2.4	4.51	4.34	3.1	3.95
8,36	1.38	3.96	3.96	0.84	1.98	4.16	4.45	3.06	4.06
8,80	2	4.18	4.18	1.39	1.64	4.12	4.79	3.13	4.27
9,24	2.7	4.51	4.51	2.03	1.45	4.13	4.54	3.13	4.58
9,68	3.47	5.49	5.49	2.75	1.54	4.22	4.15	2.28	5.55
10,12	4.39	6.08	6.08	3.37	2.02	4.51	3.91	1.74	6.11
10,56	4.65	5.46	5.46	3.49	1.57	4.09	3.94	1.87	5.49
11,00	5	5.62	5.62	3.73	0.7	3.58	4.03	2.07	5.65
11,44	6.24	6.19	6.19	4.35	0.76	3.49	4	1.41	6.17
11,88	7.28	5.62	5.62	5.11	1.82	3.45	3.92	0.69	5.54
12,32	7.1	4.83	4.83	5.12	2.54	2.58	3.52	0.29	4.84
12,76	7.02	4.63	4.63	5.28	2.57	1.78	2.8	1.2	4.73
13,20	7.41	4.69	4.69	5.8	2.58	1.4	2.21	1.97	4.98
13,64	8.15	5.01	5.01	6.86	2.83	1.57	1.64	2.59	5.77
14,08	8.82	5.25	5.25	8.26	3.47	2.08	1.18	3	6.36
14,52	9.16	5.48	5.48	9.21	4.25	2.54	0.78	3.29	6.9
14,96	8.67	5.17	5.17	9.68	5.36	3.03	0.42	3.14	6.77
15,40	7.56	4.59	4.59	8.84	6.84	3.27	0.25	2.84	6.31
15,84	6.24	3.9	3.9	7.61	8.2	3.23	0.17	2.24	5.53
16,28	4.92	3.15	3.15	6.06	8.09	3.22	1.56	2.17	4.59
16,72	3.53	2.38	2.38	4.39	7.37	3.39	1.79	1.87	3.39
17,16	2.48	1.6	1.6	3.05	6.15	3.41	1.69	1.41	2.9
17,60	2.07	1.22	1.22	2.6	4.47	2.97	2.89	1.22	2.56
18,04	3.46	0.85	0.85	1.63	3.9	2.82	0	3.09	1.16
18,48	4.54	2.79	2.79	2.63	2.87	2.14	0	2.99	0.3
18,92	3.94	2.99	2.99	1.82	1.47	1.12	0	2.35	0.38
19,36	4.37	2.45	2.45	2.11	1.69	2.67	0	2.19	1.68
19,80	0	3.17	3.17	0	1.1	1.71	0	0	0
20,24	0	0	0	0	0.66	1.64	0	0	0
20,68	0	0	0	0	0.61	2.84	0	0	0
21,12	0	0	0	0	1.17	0	0	0	0

