

6
A51

МИНИСТЕРСТВО ВЫСШЕГО И СРЕДНЕГО СПЕЦИАЛЬНОГО
ОБРАЗОВАНИЯ РСФСР

ЛЕНИНГРАДСКИЙ ОРДЕНА ЛЕНИНА ПОЛИТЕХНИЧЕСКИЙ
ИНСТИТУТ имени М. И. КАЛИНИНА

На правах рукописи

ЖЕБРОВСКАЯ
Нонна Николаевна

ИССЛЕДОВАНИЕ ИНВАРИАНТНЫХ ПРИЗНАКОВ
РЕЧЕВЫХ СИГНАЛОВ

05.13.01 — Техническая кибернетика
и теория информации

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

ЛЕНИНГРАД
1972

+

Министерство Высшего и среднего специального
образования РСФСР
Ленинградский ордене Ленина политехнический институт
имени М.И.Коллинине

На правах рукописи

КЕБРОВСКАЯ НОННА НИКОЛАЕВНА

ИССЛЕДОВАНИЕ ИНВАРИАНТНЫХ ПРИЗНАКОВ
РЕЧЕВЫХ СИГНАЛОВ

05.18.01 - техническая кибернетика
и теория информации

Автореферат диссертации на соискание ученой
степени кандидата технических наук

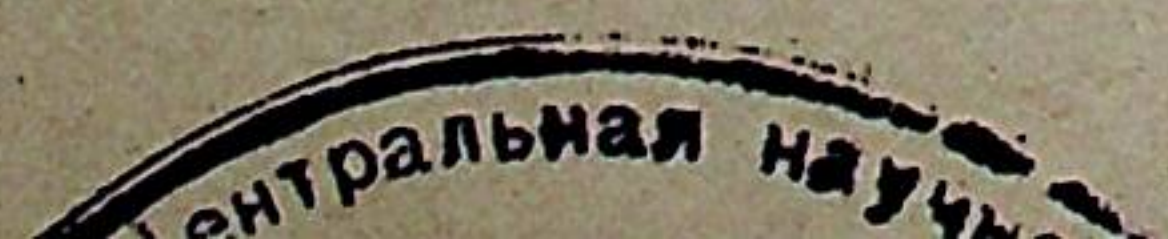
Ленинград, 1972

Работа выполнена в Ленинградском политехническом институте имени М.И.Калинина.

Научный руководитель: доцент к.т.н. ЕВМЕНОВ В.П.
Официальные оппоненты: профессор д.т.н. РОСТОВЦЕВ Ю.Г.
начальник НИЛ[№] Ленинградской Военно-инженерной Краснознаменной академии им.А.Ф.Можайского;
ст.преподаватель к.т.н. ГОРДИЕНКО Б.А. Военной ордена Ленина Краснознаменной академии связи.

Ведущее предприятие: *НИИ Радио*
Автореферат выслан "16" Июня 1973 г.
Защита диссертации состоится "16" марта 1973 г.
в " " часов на заседании Совета ФЭ Ленинградского политехнического института (194251, г. Ленинград, Политехническая улица, 29, 21-30 корпус, 224 аудитория).

Ученый секретарь Совета
доцент А.П.Хилинский



Тема диссертационной работы "Исследование инвариантных признаков речевых сигналов" касается проблемы автоматического распознавания слуховых образов и проблемы построения эффективных систем передачи речи. Несмотря на существенные успехи, достигнутые за последние годы в общей речевой проблеме, до сих пор не представляется возможным спроектировать эффективной системы автоматического распознавания слитной речи с высокой надежностью, не зависящей от индивидуальностей дикторов и работающей в реальном масштабе времени.

Коренное решение проблемы автоматического распознавания слитной речи произвольного лица тормозится по мнению многих исследователей, и мы также присоединяемся к нему, только тем, что до сих пор не найдены инвариантные признаки фонем. Под инвариантными признаками фонем понимают те или иные параметры речевого сигнала, определяющие фонемное качество данного сигнала и не зависящее от индивидуальных особенностей диктора.

Отсутствие успехов в поиске инвариантных признаков фонем не позволяет также спроектировать эффективную систему передачи речи. Наиболее эффективной, в смысле количества передаваемой информации, системой передачи речи является фонемный вокодер. Однако в настоящее время фонемные вокодеры имеют неудовлетворительную разборчивость и качество звучания. В связи с вышеизложенным очевидна актуальность темы диссертации.

В проблеме исследования инвариантных признаков речевых сигналов существенным является вопрос о критерии, который кладется в основу выбора того или иного признака или системы признаков. Считается, что таким критерием является информативность признака, под этим понимается, что данный признак должен включать в себя информацию о фонемной принадлежности. Информативность признака должна проверяться синтезом, то есть синтезированный по данному признаку сигнал должен обеспечить соответствующий эффект при восприятии, то есть сообщить сигналу фонемное качество. Поиск инвариантных признаков является своего рода минимизацией речевых сигналов. Однако

попытки глубокой минимизации, необходимой для выделения иско-
мых признаков фонем, приводят к значительному снижению раз-
борчивости и качества звучания, что объясняют нарушением фо-
немных признаков. Это заставляет исследователей увеличивать
количество информации в признаках. В результате предлагаемые
разными исследователями признаки оказывались избыточными и
недежностью распознавания по ним зависит от индивидуальностей
дикторов, интонации произношения и даже эмоционального состоя-
ния диктора. Снижение разборчивости и качества речи при ми-
нимизации целесообразно объяснить несколько иначе. Можно пред-
положить, что наряду с фонемным признаком в речевых сигналах
существуют какие-то специфические особенности, которые не
обладают информативностью, не определяют фонемного качества
сигнала, но искажение или отсутствие которых затрудняет или
делает невозможным процесс восприятия речевого сигнала, что
и фиксируется как снижение качества звучания и разборчивости.

Минимизация речевых сигналов при поиске информативных
признаков вполне может приводить к нарушению в первую очередь
предполагаемых специфических особенностей, без которых слух
не может успешно справиться с задачей выделения и "измерения"
параметров информативных признаков. Если выдвинутая гипотеза
верна, то проверка информативности признаков синтезом не
будет отражать истинное положение, если не учитывать при син-
тезе этих специфических свойств процессов восприятия, то есть
если искусственно не воспроизводить эти особенности речевых
сигналов.

Таким образом, на пути поиска инвариантных признаков рече-
вых сигналов встала задача - выяснить существуют ли в речевых
сигналах предполагаемые специфические особенности, способству-
ющие слуховому выделению фонемных признаков. Предложенная ги-
потеза определяет наш подход к решению проблемы исследования
инвариантных признаков речевых сигналов. Первый этап исследо-
вания включает в себя анализ речевых сигналов с целью поиска
предполагаемых особенностей, выявления их закономерности и

определения количественных оценок их параметров. Второй этап
исследования посвящен поиску информативных признаков с ис-
пользованием синтеза для контроля информативности, при этом
в сигналы искусственно вводятся найденные специфические осо-
бенности.

Для выполнения поставленной задачи проводился экспери-
ментальный анализ временной структуры речевых сигналов с
использованием ЦВМ БЭСМ-2М. Параллельно с исследованием вре-
менного представления сигналов в работе проводился анализ
их спектров. При этом, отказавшись от общепринятого метода
перехода к спектральному представлению преобразованием Фурье,
речевые сигналы исследовались на присутствие в них ритми-
ческих компонент методами выявления скрытых периодичностей.
Метод аппроксимации функций рядом Фурье не позволяет выявить
характерные черты исследуемого процесса, в частности выявить,
какие периодические компоненты имеются в сигнале. Действи-
тельно, периоды частотных составляющих ряда разложения не-
начаются заранее и зависят от длительности аппроксимируемого
отрезка функции. В методах же выявления скрытых периодично-
стей период искомой компоненты и подлежит определению, при
этом спектральное представление речевых сигналов существен-
но упрощается, оно описывается совокупностью ритмических
компонент, периоды, амплитуды и фазы которых определяются
этими методами.

Все эксперименты проводились на клипированной речи.
Квантованный по времени клипированный сигнал вводился
в машину в реальном масштабе времени отсчетами с шагом кван-
тования 0,0625 мсек таким образом, что в каждом разряде ячейки
записывается значение ординаты. Положительным значениям вре-
менной функции клипированного сигнала соответствуют еди-
ничные состояния разрядов, отрицательным - нулевые состоя-
ния, так что в машине клипированный сигнал представляет...
собой последовательность чередующихся групп единиц и групп
нулей. Длительность группы из единиц соответствует длитель-

ности положительной полуволны колебания, а длительность группы из нулей - длительности отрицательной полуволны речевого колебания. Ввод, а также прослушивание введенного сигнала осуществляются при выполнении машинных команд обращения к магнитной ленте (фиктивная запись и фиктивное считывание с магнитной ленты). Для прослушивания сигнала усилитель воспроизведения с громкоговорителем подключается к выходу триггера числа машины, на который, при выполнении команды фиктивной записи на магнитную ленту, последовательно считываются состояния рядов массива ячеек, отведенного под речевой сигнал. Аппаратура клипирования, ввода и воспроизведения проектировалась не в рамках данной диссертации, поэтому принципиальный охват в работе не приводится. Разборчивость введенной в ЦВМ клипированной речи составляла в среднем 96%. Максимальная длительность речевого отрезка, вводимого за один прием, составляет 4,8 сек при частоте квантования 16кГц. Для долговременного хранения образцов речевых сигналов использовались перфокарты. В наборе карточки речевых сигналов участвовало 10 дикторов (5 мужчин и 5 женщин). В состав карточки входят изолированно произнесенные гласные, слоги, слова и фразы русской речи, произносимые полным стилем. Каждый образец речевого сигнала представлен в карточке пятью реализациями (каждый диктор по пять раз произносил данный образец). В общей сложности анализу подверглось 500 реализаций речевых сигналов.

Инвариантные признаки фонем представляют собой минимизированное описание речевого сигнала, поэтому их поиск осуществлялся посредством постепенной минимизации сигнала с контролем производимого эффекта при восприятии. Поиск минимального, в смысле количества информации, временного представления сигнала, которое еще производит нужный эффект при восприятии, должен привести к такому моменту, когда дальнейшая минимизация будет сопровождаться значительным ухудшением разборчивости. Необходимо тщательно проанализировать и выяснить за счет чего ухудшается разборчивость,

разрушился ли код фонемы (инвариантный признак с минимально возможным количеством информации) или искажены особенности речевых сигналов, не несущие фонемной информации, но без которых привычные условия восприятия нарушаются. В экспериментах по минимизации речевых сигналов учитывались две гипотезы фонетического кодирования фонем - формантная и гипотеза кодирования фонетических признаков законами изменения спектральных распределений речевого сигнала. Это привело к необходимости интерпретировать спектральное представление речевых сигналов совокупностью спектральных разрезов на длительности сигнала и рассматривать эту совокупность в двух координатных плоскостях "частота-время" и "интенсивность-время" в виде двух непрерывных функций для каждой ритмической компоненты сигнала. Известно, что спектральная функция на длительности фонемы изменяется, изменяется некоторым образом положение частотных максимумов (формант) и их интенсивность. В экспериментах по минимизации исследовалось раздельно влияние на разборчивость и качество звучания временного изменения амплитуд частотных составляющих при фиксированных на всей длительности значениях частот, и временного изменения частот, при постоянстве их амплитуд. Желаемые изменения спектральной функции речевого сигнала производились соответствующим изменением временной структуры сигнала. Все манипуляции с временным представлением осуществлялись программным путем. Для каждого временного изменения исходного сигнала появилась специальная программа. Используемые алгоритмы приводятся в работе при описании соответствующих экспериментов. В общем, все эти программы включают в себя блок анализа текущей временной структуры исходного сигнала и блок внесения изменений. Синтез видоизмененных речевых сигналов сводился к "расстановке" в массиве ячеек ОЗУ групп единиц и групп нулей в нужных местах в соответствии с исходными данными и поставленной целью эксперимента. Синтезированные программным путем речевые сигналы прослушивались и производились оценки их разборчивости. Серия сигналов одного эксперимента предъявлялась группе слушателей и на основе

нии их ответов составлялось заключение о результатах экспериментов. Процент разборчивости сигналов определялся отношением числа правильных ответов к общему числу ответов и усреднялся по всем звукам данного эксперимента. Однако в общей методике экспериментального анализа наиболее важны не абсолютные оценки разборчивости, а сравнительные, поэтому определялся процент снижения или повышения разборчивости синтезируемого сигнала по отношению к сигналам других экспериментов или к исходному клиппированному. Кроме того фиксировалась сравнительная качественная характеристика звучания сигнала по отношению к исходному клиппированному сигналу, при этом слушателей инструктировали, на какие характеристики звучания им следует обратить внимание /громкость, естественность звучания, особенности интонационного рисунка, тембр, открытость звука/. Влияние тренированности аудиторов на оценки разборчивости по возможности исключалось.

В главе I описывается экспериментальный анализ временной структуры речевых сигналов, проводимый с целью минимизации и выявления предполагаемых специфических особенностей, способствующих слуховому выделению фонемных признаков. Первый этап минимизации состоял в выяснении важности для разборчивости длительностей положительных и отрицательных полуволн речевого колебания. Программным путем в клиппированных речевых сигналах все группы единичных разрядов, независимо от их длительности, заменялись одним единичным разрядом, расположенным в центре группы. Такие сигналы, состоящие из одних центров положительных полуволн колебания (в дальнейшем будем их называть сигналами, синтезированными по центрам), не теряют разборчивости. Слушатели отметили понижение громкости и незначительное ухудшение качества звучания, выражающееся в неестественности звучания. Синтезированные по центрам сигналы в отличие от исходных клиппированных имеют постоянную интенсивность полуволн колебания от периода к периоду. Это обстоятельство в спектральном представлении

проявляется как частичное нарушение естественного изменения на длительности фонемы амплитуд ритмических составляющих сигнала. Но указанная процедура минимизации не нарушает положений по оси частот периодических компонент исходного сигнала и естественного изменения основного тона на длительности сигнала.

Синтезированные по центрам речевые сигналы имеют один единственный переменный параметр - интервал времени между центрами положительных полуволн колебания. Для локализованных речевых сигналов структуру колебания на периоде основного тона можно представить набором параметров T_n - временными интервалами между соседними центрами, в дальнейшем набор параметров T_n , составляющих в сумме длительность периода основного тона, будем называть маской. Временная структура гласных звуков характеризуется тем, что на его длительности наблюдается изменение маски. В небольших пределах от периода к периоду изменяются значения параметров T_n , причем в соответствии с изменением периода основного тона. Изменяется также и число параметров T_n в маске, создавая неоднородность временной структуры звука. Однако можно выделить небольшое число типов масок, отличающихся числом параметров T_n . Следующий этап минимизации состоял в выяснении возможности устранения неоднородности временной структуры локализованных звуков. Это достигалось синтезом однородного сигнала с каждой выделенной маской. Временная структура сигнала приводилась к однородной (на всех периодах маска была одинакова), но полностью сохранялись особенности изменения основного тона и параметров T_n . По результатам прослушивания отбирались те маски, для которых был наибольший процент идентификации с главным звуком исходной реализации. Разборчивость однородных звуков снижалась по сравнению с разборчивостью неоднородных, синтезированных по центрам, в среднем на 8%. Все аудиторы отмечали неестественность звучания. В синтезированных по маскам однородных сигналах полностью отсутствует естественное изменение амплитуд периодических составляющих сигнала, но не

нарушается изменение основного тона и положение по оси частот периодических компонент на длительности звука. Результат этого эксперимента позволяет считать, что маска и временная динамика параметров T_k включает в себя информацию о фонемной принадлежности сигнала, достаточную для различения фонемы.

Следующий этап минимизации касается уже временного изменения основного тона. Цель состояла в выяснении роли при восприятии изменения периодов основного тона и параметров T_k на длительности сигнала, имеют ли они отношение к фонемной принадлежности сигнала. В однородных сигналах, синтезированных по центрам, устранилось временное изменение параметров T_k и периодов основного тона. Выравнивание периодов основного тона производилось пропорциональным увеличением или уменьшением значений T_k внутри периода. Такие монотонные главные фонемы воспринимались как сложный тон (автомобильный гудок), аудиторы не относили их к речеподобным. В этом случае имеет смысл говорить о полной потере разборчивости и фонемного качества. Потеря фонемного качества могла бы трактоваться так, что временное изменение параметров T_k и периодов основного тона существенно для фонемного различия главных. Однако, учитывая выказанное предположение о том, что в речевых сигналах могут быть специфические особенности не информативные, но нарушение которых тоже приводит к потере фонемного смысла, следует воздержаться от такой трактовки, тем более, что она противоречит практике вокодерной передачи речи с использованием автономного генератора основного тона. Скорее всего фонемные признаки в сигналах последнего эксперимента не нарушены, но почему-то не выделяются при слуховой обработке сигнала, что и проявляется как потеря фонемного смысла. Цель следующего четвертого эксперимента состояла в доказательстве, что фонемные признаки в сигналах предыдущего эксперимента сохранены, а потеря фонемного смысла происходит по причине нарушения специфической особенности, способствующей слуховому выделению фонемных признаков. Осуществлялся перенос временного изменения основного тона и

параметров T_k с одной фонемы на другую. Синтезировался звук с маской, принадлежащей одной фонеме, и относительными изменениями периодов основного тона и параметров T_k , перенятыми с другой фонемы. Прослушивание показало, что сигналы приобретали фонемное качество звука, у которого была взята маска при этом разборчивость имела такие же оценки как для однородных, синтезированных по центрам, сигналов. Этот эксперимент подтверждает предположение о том, что минимизация, последовательно проведенная в указанных трех экспериментах и приводящая к потере фонемного качества, все же не приводит к нарушению фонемных признаков. Фонемное качество восстанавливается при введении в сигнал неустойчивости частоты основного тона, взятой от любого речевого сигнала. Временное изменение основного тона очевидно включает в себя специфическую особенность, способствующую слуховому выделению фонемных признаков. Следующий эксперимент ставился для выяснения кажущегося противоречия между результатом третьего эксперимента и практикой синтеза речи с использованием автономного генератора основного тона. Сигналы третьего эксперимента отличаются только тем, что в них отсутствует естественная динамика амплитуд периодических компонент на длительности сигнала. Выяснялась роль изменения амплитуд периодических компонент сигнала при сохранении изменения частот периодических компонент на длительности сигнала. В противоположность эксперименту 2, где фиксировались постоянными амплитуды периодических компонент, в исходных клипированных сигналах выравнивались периоды основного тона пропорциональным увеличением или уменьшением длительностей полуволн речевого колебания, при этом естественное изменение амплитуд сохранялось. Полученные сигналы аналогичны сигналам, синтезируемым в вокодерах с автономным генератором основного тона, разница только в клипировании. Разборчивость по сравнению с сигналами, синтезированными по центрам снижена в среднем на 4%. Отмечалась неестественность звучания. Результаты двух последних экспериментов приводят к выводу, что фонемное качество сигналов эксперимента 8 восстанавливается либо при наличии

временного изменения частот периодических составляющих сигнала, если их амплитуды постоянны, либо при наличии временного изменения амплитуд, если частоты составляющих постоянны.

Результаты изложенных экспериментов по исследованию временной структуры вокализованных звуков речи дают право сделать следующие выводы:

1. Изменение на длительности фонемы частоты основного тона, приводящее к соответствующему изменению частот периодических компонент сигнала, и временное изменение амплитуд периодических компонент содержит не информативные специфические особенности, способствующие слуховому выделению фонемных признаков.

2. В связи с подтверждением гипотезы о существовании в речевых сигналах указанных особенностей, процесс восприятия вокализованных фонем можно рассматривать состоящим из двух моментов: первый - для правильного слухового распознавания фонемы в сигнале должен быть сохранен код фонемы, который можно описать или спектральными признаками или маской во временном представлении; второй - для того, чтобы код фонемы создавал фонемное качество сигнала при слуховом восприятии, в нем должны быть сохранены специфические особенности, состоящие во временном изменении параметров кода на длительности фонемы.

Слуховое выделение фонемных признаков может осуществляться при наличии в сигнале хотя бы одного условия: либо - изменения частоты основного тона и соответствующего ему изменения частот периодических компонент сигнала, либо - изменения их амплитуд на длительности сигнала. Хорошая разборчивость и качество звучания обеспечивается, если в сигнале есть эти два условия "слухового измерения" фонемных признаков. При отсутствии одного из них снижается качество и разборчивость, а при отсутствии обоих условий - фонемные признаки не выделяются. Это проявляется как полная потеря разборчивости и фонемного качества. Образно говоря, специфические особенности, состоящие во временной нестабильности параметров фонемных признаков, являются "транспортным средством" для выявления

кода фонемы в слуховом анализаторе.

П глава посвящена определению количественных оценок параметров составляющей временного изменения частоты основного тона, которая способствует слуховому выделению фонемных признаков. Известно, что основной тон характеризуется плавным изменением /мелодия основного тона/, которое создает при восприятии интонацию произношения, и иррегулярным, придающим живость звучания, натуральность, и являющимся индивидуальной особенностью диктора. Экспериментально было установлено, что из этих двух составляющих слуховому выделению фонемных признаков способствует плавное изменение частоты основного тона. Проводился анализ для выяснения, является ли этим фактором интонационная мелодия тона, или в ней присутствует специфическая составляющая. Для выполнения этого исследования был набран большой статистический материал кривых временного изменения периодов основного тона вокализованных звуков и участков вокализации из слов и фраз.

Кривые изменения периодов основного тона клиппированных речевых сигналов определялись на ЦВМ по кратковременной функции автокорреляции. В работе приводится описание этого алгоритма и результаты исследования по выбору весовой функции и возможности алгоритма фиксировать быстрые изменения периодов основного тона.

Автокорреляционная функция вычисляется по формуле:

$$R(n, lk) = \frac{1}{A \cdot m} \sum_{j=lk}^{lk+m} \langle A_j^0 \rangle \oplus \langle A_j^n \rangle$$

- где $\langle A_j^0 \rangle$ - массив исходного клиппированного сигнала,
- $\langle A_j^n \rangle$ - массив сигнала, сдвинутого на n разрядов,
- A - число разрядов ячейки машины (для БЭСМ-2М $A=39$),
- l - шаг времени, выраженный числом ячеек,
- m - длительность весовой функции, выраженная числом ячеек (весов. функция берется в виде прямоугольного окна),
- K - шаг сдвига временного окна по массиву сигнала.

Выделяемым периодом является интервал времени между первым и вторым максимумом функции автокорреляции. Разрешающая способность алгоритма по времени исследовалась на синтезированных для этой цели функциях, учитывающих структуру вокализованных речевых сигналов. Исследование показало, что при длительности весовой функции в один период $T = 2/\omega$ алгоритм выделяет каждый период функции, а при длительности весовой функции, равной 3-4 периодам основного тона, выделяется только плавная составляющая заданного изменения периодов.

Анализ кривых /анализу подвергалось около 150 реализаций/ монотонного изменения периодов основного тона показал, что для изолированно произнесенных гласных звуков наблюдается колебательный характер изменения основного тона, близкий к гармоническому. В некоторых реализациях можно проследить 1,5-2 периода медленного колебания периодов основного тона относительно среднего значения. Колебание гармонического характера особенно ярко выражено на начальных участках звуков, а для некоторых реализаций и на всей длительности. Многие реализации имеют затухающие к концу звука колебания. Однако для участков вокализации слитной речи колебание периодов основного тона носит на первый взгляд не гармонический характер. Различие характера колебания основного тона изолированных гласных и участков вокализации слитной речи свидетельствует о том, что искомый фактор представлен отдельной составляющей монотонного изменения основного тона. Монотонное изменение основного тона можно рассматривать состоящим из двух составляющих: плавного изменения основного тона, определяющего интонацию, и медленного почти гармонического колебания основного тона относительно первой составляющей. Первая составляющая, ответственная за интонационную окраску, представляет собой медленное изменение среднего значения периода основного тона. Поскольку для изолированно произнесенных гласных интонация отсутствует (среднее значение периода основного тона постоянно на всей длительности), то вторая составляющая видна без нарушения формы колебания. Для участков вокализации слитной речи

характерна интонационная мелодия, и за счет изменения среднего значения периода основного тона вторая составляющая проявляется с нарушением форм колебания. Если участки вокализации разбить на сегменты, в пределах которых среднее значение периода основного тона не изменяется, то окажется, что значения периодов основного тона составляют гармоническую зависимость. Медленное колебание основного тона может быть выделено из кривой изменения основного тона, если его считать результатом угловой модуляции сигнала некоторой модулирующей функцией. В первом приближении можно аппроксимировать искомую составляющую изменения периодов основного тона гармонической зависимостью. Тогда в терминах радиотехники средняя частота основного тона является несущей частотой, которая изменяется во времени по гармоническому закону модуляции. Для определения оценок параметров медленного колебания определялся вид угловой модуляции речевых сигналов (фазовая или частотная), а также параметры модулирующей функции. Изменение во времени периодов основного тона в случае гармонической угловой модуляции описывается тремя параметрами: τ_{cp} , $\Delta\tau$, N или Ω

- τ_{cp} - среднее значение периода основного тона,
- $\Delta\tau$ - максимальное отклонение от среднего значения,
- N - период модуляции,
- Ω - частота модуляции.

По кривым временного изменения основного тона определялись значения параметров τ_{cp} , $\Delta\tau$, Ω , они оказались независимыми друг от друга. В полученном статистическом материале диапазон изменения τ_{cp} соответствует по частоте 150-400 гц. Значение параметра $\Delta\tau$ лежит в диапазоне 0,1-0,4 мсек. Частота модуляции для всего диапазона основного тона не выходит за пределы 5 рад/сек - 40 рад/сек (0,9-6 гц) при нормальном состоянии дикторов. При небольшой предшествующей произношению, физической нагрузке, состоящей в подъеме и спуске по четырем ступеням, диапазон частоты модуляции расширялся до 80 рад/сек (12 гц). Эксперимент по определению связи между частотой модуляции и физической нагрузкой был предпринят для того, чтобы на большем диапазоне изменения аргумента проследить зависи-

мости индекса модуляции и девиации частоты от частоты модуляции, по которым можно отличить фазовую и частотную модуляции.

Для определения вида угловой модуляции речевых сигналов были выведены аналитические зависимости, связывающие параметры угловой модуляции - индекс модуляции и девиацию частоты - с параметрами изменения периодов основного тона:

$$\Theta = \frac{\pi \cdot \Delta n}{n_{cp} \cdot \sin \frac{\pi}{N} (n_{cp} + \Delta n)} \quad (3)$$

$$\omega_g = \frac{2\pi^2 \cdot \Delta n}{N \cdot n_{cp} \cdot \sin \frac{\pi}{N} (n_{cp} + \Delta n)} \quad (4)$$

По этим зависимостям определялись экспериментальные значения параметров модуляции частоты основного тона вокализованных звуков речи. Для определения вида угловой модуляции строились зависимости Θ и ω_g от Ω . Поскольку оказалось, что Δn не зависит от n_{cp} , то согласно выражений (3,4) параметры модуляции зависят от среднего значения периода основного тона. Для того, чтобы использовать полный объем экспериментальных значений Θ и ω_g , независимо от среднего значения периода основного тона, параметры модуляции определялись приведенными к фиксированному среднему значению основного тона. Процедура приведения возможна, так как Δn и Ω не зависят от n_{cp} . Зависимости приведенных параметров модуляции от частоты модуляции оказались характерными для частотной модуляции. Эксперименты по синтезу гласных звуков показали, что разборчивость сигналов, синтезированных по центрам, с введением частотной гармонической модуляции повышается в среднем на 4% по сравнению с аналогичными сигналами, но с естественным временным изменением периодов основного тона. Факт повышения разборчивости можно объяснить искусственным "исправлением" модуляционного процесса речевых сигналов. Отклонение соотношения параметров ω_g и n_{cp} от выведенной зависимости вызывало значительное ухудшение разборчивости. Экспериментально было также установлено, что модуляционный процесс, если его парамет-

ры находятся в соотношениях согласно выведенной зависимости, не фиксируется слухом как колебание высоты тона. По результатам исследования II главы можно сделать следующие выводы:

1. Временное изменение основного тона вокализованных звуков речи содержит специфическую составляющую, способствующую слуховому выделению фонемных признаков, в виде медленного колебания частоты основного тона относительно среднего значения. Частота этих колебаний не превышает 0,9-6 гц и увеличивается до 12 гц при предшествующей проанализированной небольшой физической нагрузке.

2. Медленное колебание частоты основного тона является результатом частотной модуляции, имеющей место в речевых сигналах. Параметры модуляции, характеризующие медленное колебание, зависят от среднего значения частоты тона. Девиация частоты зависит от среднего значения периода основного тона приблизительно как $1/n_{cp}$ (точная зависимость описывается формулой 4) и не зависит от частоты модуляции для данного значения частоты основного тона. Индекс модуляции является функцией двух аргументов n_{cp} и Ω .

В главе III исследуется модель параметрической системы передачи речи. Моделирование вокодера для передачи клипированной речи производилось с целью определения возможностей применения в системах компрессии речи полученных результатов о медленных колебаниях частоты основного тона. В анализируемом устройстве модели из всех присутствующих в мгновенном спектре сигнала формантных областей выбирается форманта с максимальной энергией, ее параметры F_{max} и ΔF_{max} , а также суммарная мощность сигнала вычислялись по параметрам распределения

T_{kl} - временных интервалов между центрами положительных полуволи речевого колебания. Сигнал тон/ш/м и текущие значения периодов основного тона определялись по кратковременной функции автокорреляции. Выделенные из речи параметры передаются (в условиях моделирования на машине - запоминаются) отсчетами через каждые 25мсек. Разборчивость речи на выходе такой системы незначительно ухудшается по сравнению с клипированной речью на входе. Если передавать сведения о двух формантах спек-

ральной функции речевого сигнала, естественно ожидать улучшения разборчивости. Однако, не преследовалась цель рекомендации этой модели, она необходима для постановки экспериментов с каналом основного тона. Проводились эксперименты по передаче речи при нескольких режимах работы канала возбуждения модели. При режиме работы с генератором основного тона, управляемом выделенными значениями периодов основного тона, разборчивость фонем составляла в среднем 80%. При режиме работы с автономным генератором основного тона, управляемом от датчика случайных чисел, разборчивость составляла в среднем 72%. При режиме работы с автономным генератором основного тона, управляемом по гармоническому закону частотной модуляции, разборчивость составляла также 80%. Управление генератором тона осуществлялось в этом случае в соответствии с найденными оценками параметров модуляции. Для назначенного среднего значения периода основного тона девиация частоты определялась по формуле (4) при $\Delta n = 0,3$ мсек (это среднее значение параметра по всей выборке). Индекс модуляции вычисляется по формуле $\theta = \omega_g \Omega$, а Ω выбиралась произвольной из диапазона 0,9-6 гц. При режиме работы с генератором основного тона, управляемым средним за некоторый интервал значением периода основного тона, модулированного по гармонической частотной модуляции, разборчивость составляла в среднем 82%. Полученные оценки разборчивости дают право сделать вывод, что для систем компрессии речи, где не требуется передачи индивидуальных особенностей диктора, введение на приемной стороне гармонической частотной модуляции дает улучшение качества и разборчивости речи в среднем на 8%. В случае, когда необходима передача индивидуальных особенностей диктора введение частотной модуляции дает возможность уменьшить частоту передачи основного тона при сохранении разборчивости и качества, то есть повысить эффективность системы.

Глава IV посвящена исследованию инвариантных признаков гласных звуков. Эксперименты с временной структурой гласных звуков привели к предположению, что коды (фонемные признаки) можно описать небольшим числом периодических составляющих сиг-

нала, и что данное описание будет инвариантным фактору диктора. Методом выявления скрытых периодичностей был получен статистический материал по частному составу (периодограммы) гласных звуков, изолированно произнесенных и вырезанных из слогов, слов и фраз. Анализ частотного состава одинаковых гласных разных дикторов показал, что значения частот периодических составляющих сигналов, имеют значительный разброс (анализировались частоты, определяющие первую и вторую формантные области спектра). В то же время, если определять частоты периодических компонент сигнала номерами гармоник основного тона, то оказалось, что существуют вполне определенные зависимости между номерами гармоник, определяющими первую и вторую форманты, и средним значением периода основного тона для каждого гласного звука. В результате обработки в указанном выше смысле статистического материала по частотному составу гласных звуков были получены зависимости номеров гармоник, определяющих первую и вторую форманты, от среднего значения периода основного тона для каждого гласного звука. Эти графики приведены на рис. I, где сплошной линией показано изменение в зависимости от периода основного тона номера гармоники, определяющего первую форманту, а пунктиром - изменение в зависимости от периода основного тона номера гармоники, определяющего вторую форманту. Как следует из графиков, указанные зависимости линейные и для разных гласных отличаются крутизной. По данным графиков отстроились фонемные области в координатах номеров гармоник, определяющих первую и вторую форманты, области не пересекаются. Информативность выявленных фонемных признаков проверялась синтезом. Данными для синтеза были среднее значение периода основного тона и номера гармоник для соответствующего гласного звука. Синтез осуществлялся программным путем суммированием мгновенных значений гармонических составляющих с последующим клиппированием сигнала. Период основного тона изменялся по закону гармонической частотной модуляции, при этом текущее значение периода основного тона определялось решением уравнения:

$$\sin\left[\frac{2\pi}{n_f}(n_i + k_0) - \theta \cdot \sin \frac{2\pi}{N}(n_i + k_0)\right] = 0 \quad (5)$$

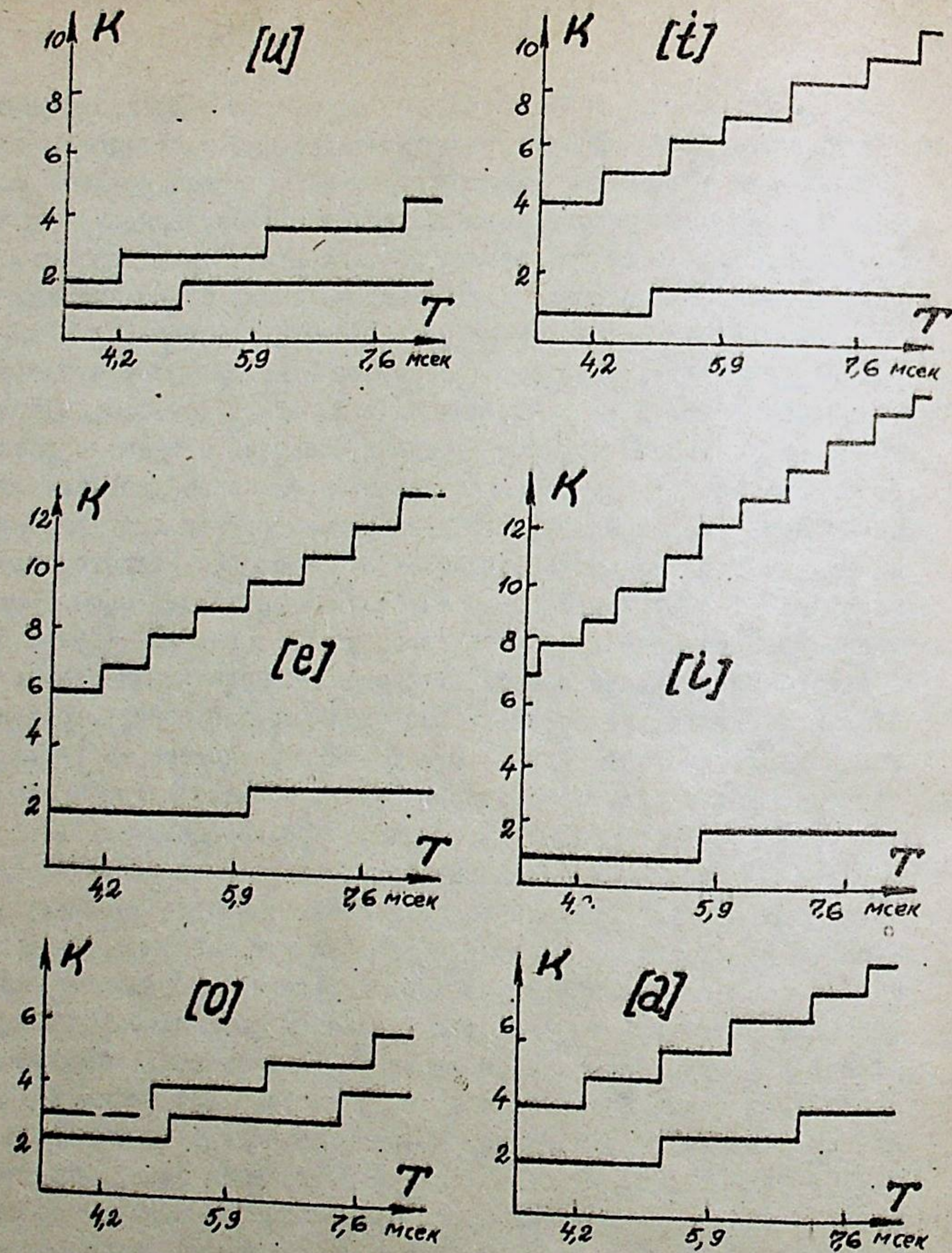


Рис. I. Зависимости номеров гармоник, определяющих I и II формантные области, от среднего значения периода основного тона для шести гласных.

где n_i - текущее значение периода основного тона;
 K_0 - начальная фаза модулирующего колебания (выбирается произвольной),

а Θ - индекс модуляции вычислялся для ΔT и $\Delta T = 0,3$ мсек по формуле (3).

Разборчивость гласных, синтезированных по центрам с введенной гармонической частотной модуляцией не отличалась от разборчивости гласных, с естественным изменением основного тона. Таким образом, в результате проведенного анализа были определены коды гласных звуков, представленные небольшим числом гармоник основного тона, номера которых зависят от периода основного тона и фонемной принадлежности гласного.

Результаты выполненного исследования состоят в следующем:

1. На длительности изолированных гласных, а также на вокализованных участках слитной речи обнаружены медленные, почти гармонические колебания частоты основного тона. Частота этого колебания лежит в диапазоне 0,9-6 гц при нормальном состоянии диктора и увеличивается до 12 гц предшествующей произношению небольшой физической нагрузке. Зависимости между частотой этого колебания и средним значением частоты основного тона не обнаружено. Амплитуда колебания периода составляет в среднем 0,3 мсек и не зависит от среднего значения частоты основного тона.

2. Показано, что колебание частоты основного тона можно считать результатом частотной модуляции. Небольшая зависимость параметров модуляции от среднего значения частоты основного тона, которой подчиняется модуляционный процесс в речевых сигналах.

3. Определена роль колебания частоты основного тона для процесса восприятия речевых вокализованных сигналов. Ритмическое качение во времени частоты основного тона и вызванное им качение частот периодических компонент сигнала является одним из условий слухового выделения фонемных признаков вокализованных звуков речи.

4. Сделан вывод, что вторым условием слухового выделения фонемных признаков является аналогичное колебание во времени амплитуд периодических составляющих сигнала. Количественной оценки параметров их изменения в работе не проводилось.

5. Экспериментально доказано, что качание периодических составляющих сигнала в плоскости "частота-время" и качание периодических составляющих сигнала в плоскости "интенсивность-время" функционально взаимозаменяемы. При отсутствии в речевых сигналах одного из условий слухового выделения фонемных признаков ухудшается качество звучания и снижается разборчивость. При отсутствии в речевых сигналах обоих условий слухового выделения фонемных признаков фонемный смысл речевого сигнала не может быть обнаружен (звук воспринимается как сложный тон, несмотря на присутствие в сигнале фонемного признака).

6. Явление частотной модуляции речевых сигналов до сих пор ускользало от исследователей по причине малой амплитуды качания частоты основного тона и недостаточной точности анализа речевых сигналов, а также по причине маскировки этого колебания интонационной кривой основного тона. Монотонное изменение частоты основного тона принималось за интонационную мелодию. Как оказалось, в функциональном отношении монотонное изменение основного тона несет две нагрузки; создает эффект интонации и способствует слуховому выделению фонемных признаков, при этом вторую нагрузку выполняет выявленная составляющая монотонного изменения, и колебание частоты основного тона происходит относительно кривой интонационности, которую можно рассматривать как изменение среднего значения частоты основного тона. При этом в зависимости от изменения среднего значения частоты основного тона изменяются и параметры колебания основного тона - девиация частоты и индекс модуляции. Модуляционный процесс в речевых сигналах не фиксируется слухом как колебание высоты тона.

7. Обнаруженное свойство слухового восприятия локализованных звуков речи позволило провести минимизацию временного и частотного представления и определить инвариантные признаки гласных фонем, которые можно описать или масками во времен-

ном представлении (набор временных интервалов между центрами положительных полуволн речевого колебания), или номерами гармоник основного тона. Оказалось, что параметры фонемных признаков - номера гармоник или параметры T_m маски - зависят от частоты основного тона диктора. Экспериментально определены зависимости номеров гармоник, определяющих первую и вторую формантные области, от среднего значения периода основного тона для шести гласных русской речи.

8. Экспериментами на модели вокодера клиппированной речи доказано, что результаты исследования, касающиеся количественной оценки параметров модуляционного процесса в речевых сигналах, имеют большое практическое значение для построения эффективных систем передачи речи. Установлено, что введение частотной модуляции основного тона с гармонической модулирующей функцией повышает качество и разборчивость речи в системах передачи речи с автономным генератором основного тона.

По теме диссертации опубликовано:

- I. Временное изменение основного тона гласных фонем. Труды ЛПИ, № 291, Теория и техника вычислительных устройств, 1968, 141-146 (в соавторстве).

РТИ тип.ВМР зак.806 тир.140 М-60807 15.12.72