

Выпуск 578

Р. В. ГУДОНАВИЧЮС, П. П. КЕМЕШИС,
А. Б. ЧИТАВИЧЮС

**РАСПОЗНАВАНИЕ
РЕЧЕВЫХ СИГНАЛОВ
ПО ИХ СТРУКТУРНЫМ
СВОЙСТВАМ**



«ЭНЕРГИЯ»
ЛЕНИНГРАД 1977

6Ф6.5
Г93

УДК 681.3.01

Редакционная коллегия:

И. В. Антик, Г. Т. Артамонов, А. И. Бертинов, М. А. Бояр-
ченков, А. А. Воронов, Л. М. Закс, В. С. Малов, В. Э. Низе,
Д. А. Поспелов, И. В. Прангишвили, О. В. Слежановский,
Ф. Э. Темников, М. Г. Чиликин, А. С. Шаталов

Рецензент Л. Л. Мясников

Гудонавичюс Р. В., Кемешис П. П., Читави-
чюс А. Б.

Г93 Распознавание речевых сигналов по их струк-
турным свойствам. Л., «Энергия», 1977.

64 с. с ил. (Б-ка по автоматике. Вып. 578).

Книга является первой попыткой систематического рассмотре-
ния теоретических и практических вопросов использования струк-
турных свойств речевых сигналов для их распознавания. Рассматри-
ваются основы построения устройств выделения признаков по струк-
турным свойствам сигналов, приводятся их конкретные схемы, ре-
зультаты экспериментов по распознаванию устных команд.

Книга может быть полезна инженерно-техническим и научным
работникам, занимающимся вопросами обработки информации в ав-
томатических системах управления, разрабатывающим аппаратуру
для обработки и распознавания сигналов, а также студентам стар-
ших курсов и аспирантам соответствующих специальностей.

Г $\frac{30501-104}{051(01)-77}$ 117.76

6Ф6.5

ПРЕДИСЛОВИЕ

За последние годы появилась обширная литература, посвященная распознаванию звуковых образов. Этой проблематикой занимаются сейчас специалисты, работающие в самых различных областях: психологи, лингвисты, физиологи, математики, инженеры; интерес к ней легко объяснить, так как речь идет о проблеме, успешное решение которой позволит осуществить частичную замену интеллектуальной деятельности человека действием автоматов. Особенно остро ощущается необходимость решения этой проблемы в ряде областей технической кибернетики и прежде всего в области создания средств обмена информацией между человеком и ЭВМ.

Постановка вопроса в разных работах по распознаванию речевых образов, методика научного исследования и практические выводы, к которым приходят разные авторы, часто имеют довольно мало общего. Единого подхода к решению разнообразных задач, возникающих при выделении признаков, создании алгоритмов и синтезе механизмов, предназначенных для распознавания, до сих пор нет. Кроме того, незаслуженно мало внимания уделяется вопросам выбора первичного описания речевых сигналов. Многие исследователи [4, 5, 6] придерживаются мнения, что почти любое логически построенное первичное описание, сохраняющее информацию, может привести в конечном счете к положительным результатам при соответствующей организации вышних уровней распознающей системы. Однако такой подход не является совсем правильным, хотя бы потому что может привести к слишком сложным алгоритмам функционирования системы в целом. Наоборот, удачно выбранное первичное описание безусловно приводит к большим алгоритмическим и аппаратурным упрощениям.

Авторы считают, что существенным шагом вперед в деле решения проблемы распознавания речевых образов в целом и в деле выбора первичного их описания, в частности, может явиться использование структурной модели сигналов, представляющей собой совокупность нескольких обычно применяемых контурных моделей [13 23]. Это предположение базируется на том факте, что, в отличие от контурной модели, под которой понимают как сам сигнал, так и любые его функциональные преобразования, структурная модель позволяет создавать алгоритмы, результаты применения которых инвариантны к определенным свойствам сигналов; особенно остро эти преимущества выявляются при бедной априорной информации об исходном сигнале.

Структурные свойства сигналов уже успешно используются для решения целого ряда задач в связи, в технике измерений,

в диагностике процессов [13, 23]. Первые же опыты по распознаванию речевых образов на основе структурных свойств дали обнадеживающие результаты; дальнейшие работы сняли сомнения относительно перспективности этого направления [8, 9].

В книге, отражающей опыт авторов, накопленный в течение последних лет, рассматриваются теоретические вопросы использования структурных свойств сигналов для распознавания речевых сигналов, даются рекомендации для построения распознавателей и приводятся результаты экспериментальных исследований.

Первая глава посвящена теоретическим вопросам построения структурной модели сигналов. Во второй главе рассматриваются численные характеристики структурной модели речевых сигналов, т. е. признаки для распознавания. Третья и четвертая главы посвящены вопросам выделения признаков по структурной модели и их практического использования для распознавания.

Авторы далеки от мысли, что теперь уже можно дать законченные рецепты по распознаванию речевых сигналов; работа в этом направлении только начинается. Однако, по глубокому убеждению авторов, использование структурных свойств сигналов позволит привести работы по распознаванию речевых сигналов к состоянию, при котором распознавание будет осуществляться с высокой степенью надежности с помощью относительно простых устройств.

Авторы выражают глубокую признательность А. М. Заездному за ценные советы и помощь, оказанную им при подготовке книги.

Замечания и пожелания по книге просьба направлять по адресу: 192041, Ленинград, Марсово поле, д. 1, Ленинградское отделение издательства «Энергия».

Авторы

Глава первая

СТРУКТУРНАЯ МОДЕЛЬ РЕЧЕВОГО СИГНАЛА

1. Структурные свойства и структурные связи

Как уже отмечалось, структурная модель речевого сигнала противопоставляется привычной контурной модели, под которой понимаем либо сам сигнал, либо его функциональные преобразования. Простейшей контурной моделью сигнала является его «временная картина» (осциллограмма) $s(t)$, представляющая собой зависимость звукового давления, преобразованного в электрическую величину s , от времени t .

В равной степени контурными моделями являются функциональные преобразования исходного сигнала, т. е. модели вида $L[s(t)] = Ls$, где L — символизирует операцию преобразования сигнала $s(t)$. В частности, контурными моделями являются производные p -го порядка $s^{(p)}(t)$ или интегралы p -го порядка $s^{(-p)}(t)$ от исходного сигнала

$$s^{(p)}(t) = \frac{d^p [s(t)]}{dt^p}, \quad s^{(-p)}(t) = \int \dots \int s(t) dt.$$

Интегральные преобразования «с весом» также являются контурными моделями: к ним относится и спектральная функция $S(j\omega)$, определяемая интегральным преобразованием Фурье:

$$S(j\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt.$$

Структурная модель сигнала представляет собой определенную связь между взятыми в совпадающие или несовпадающие моменты времени значениями преобразованной (в общем случае) функции, описывающей сигнал, и значениями ее производных и интегралов, иными словами, между двумя или большим числом контурных моделей. Эту связь мы в дальнейшем будем называть структурной связью или структурной моделью, а свойства, которые она характеризует, структурными свойствами. Структурная модель — это изображение сигнала в некотором новом функциональном пространстве; структурные свойства — свойства этого изображения, характеризующие исходный сигнал.

По признаку зависимости от времени структурные связи можно разделить на две основные группы: 1) структурные связи с явным временем, 2) структурные связи с неявным временем.

Обобщенная аналитическая запись структурных связей с явным временем (обозначенных через $\xi(t)$) имеет следующий вид:

$$\xi(t) = \Phi \{L_1[s(t)], L_2[s(t)], \dots\} = \Phi(L_1s, L_2s, \dots). \quad (1)$$

Здесь $L_1[s(t)] = x_1(t)$, $L_2[s(t)] = x_2(t)$, ... — контурные модели сигнала. Символы L_1, L_2, \dots означают в общем случае сложные линейные и нелинейные трансформации сигнала, включая дифференцирование и интегрирование; Φ — определяет связь между контурными моделями. Примером может служить структурная связь вида $\xi(t) = x_1(t)/x_2(t) = \dot{s}(t)/\dot{s}(t)$. Очевидно, в этом случае L_2 символизирует операцию дифференцирования, L_1 — двойного дифференцирования, а Φ означает операцию деления.

Структурные связи с неявным временем строятся на основе использования связей с явным временем путем образования так называемых фазовых изображений в общем случае в N -мерном фазовом пространстве X, Y, Z, \dots , где X, Y, Z, \dots — прямоугольные декартовы координаты (в дальнейшем называемые фазовыми координатами). Именно в качестве фазовых координат и используются связи с явным временем, определенные выражением (1). Тогда связь с неявным временем — это изображение в фазовом пространстве X, Y, Z, \dots , где $X = \xi_1(t)$, $Y = \xi_2(t)$, $Z = \xi_3(t)$ и т. д.

Однако использование многомерного фазового пространства, по крайней мере на данном этапе исследований, вряд ли может считаться рациональным с точки зрения аппаратных затрат. Конечно, это не относится к теоретической стороне вопроса. Поэтому в дальнейшем в основном речь пойдет о двумерных фазовых изображениях, т. е. об изображениях на плоскости X, Y , за исключением некоторых теоретических построений.

Примером элементарного фазового изображения на плоскости служит изображение вида $X = \xi_1(t) = s(t)$, $Y = \xi_2(t) = \dot{s}(t)$ (или короче s, \dot{s}), хорошо известное в теории колебаний. Более сложные изображения могут иметь вид $\ddot{s}/s, \dot{s}/s; \ddot{s}/\dot{s}, \ddot{s}s/\dot{s}^2$ и т. д.

Информационные свойства структурной модели в общем смысле зависят от нее самой, т. е. от ее «конструкции». Конечно, в структурной модели не может содержаться больше информации, чем ее содержалось в исходной контурной модели, если только она была полной по количеству информации. Однако структурная модель может содержать и больше информации о сигнале, чем каждая из входящих в нее контурных моделей, если последние являются неполными моделями, отображающими какие-либо независимые свойства моделируемого сигнала. Если хотя бы одна из входящих контурных моделей является полной, то структурная модель обладает свойством математической избыточности. Отсюда следует, что некоторая информация о сигнале «повторяется» в неполных контурных моделях, образующих вместе с полной единую структурную модель. Смысл введения такой избыточности или «повторяемости» заключается в том, что при надлежащем выборе преобразования Φ в формуле (1) модели, обладающие математической избыточностью, позволяют вскрыть ряд признаков, характеризующих сигнал, которые затруднительно вскрыть при использовании только той или иной контурной модели. Вместе с тем информация, содержащаяся

в структурной модели, может оказаться более доступной для ее извлечения по простоте реализации устройств, по точности их работы, по разрешающей способности и т. д. Выбором подходящей структурной связи можно освободиться от избыточной информации или от влияния непостоянства безыформационных параметров, т. е. достичь некоторой инвариантности относительно них, а полезную информацию сделать более контрастной, более выразительной.

Так, например, нужно освободиться от влияния коэффициента усиления сигнала $s(t)$ (имеется в виду постоянный коэффициент усиления). Для этого достаточно образовать отношение первой производной к самому сигналу, т. е. достаточно рассматривать структурную связь вида $\delta(t) = \frac{\dot{s}(t)}{s(t)}$ или, например,

структурную связь вида $\lambda(t) = \frac{\ddot{s}(t) s(t)}{\dot{s}^2(t)}$, тождественно равную

единице для сигналов вида $s(t) = Ae^{\alpha t}$ т. е. инвариантную к значениям и знакам как A , так и α .

Следует отметить одну важную особенность фазовых изображений. Любое фазовое изображение представляет речевой сигнал в виде некоторой плоскостной или пространственной кривой, что само по себе является более «компактным» способом представления сигнала, чем представление в виде функции с явным временем. Кроме того, при дальнейшей обработке таких изображений удобно и выгодно привлечь идеи дифференциальной геометрии, топологии и качественной теории дифференциальных уравнений, что дает возможность интересных общетеоретических исследований и полезных практических разработок. Об этом речь пойдет в последующих параграфах и главах.

Таким образом, согласно приведенным определениям, формально можно построить бесконечное количество структурных моделей одного и того же сигнала. Однако любая математическая модель сигнала должна соответствовать конкретным условиям решаемых задач. Иными словами, особенности математического представления сигнала должны соответствовать объему и характеру априорной информации о самом сигнале. В связи с этим возникает вопрос: какие структурные связи следует использовать в задачах, связанных с распознаванием речевых сигналов? Для удобства изложения материала этот вопрос разделим на три вопроса. Первый из них связан с выбором контурных моделей, получаемых на основе предварительных преобразований речевого сигнала, т. е. с выбором трансформаций $L_1, L_2 \dots$ в выражении (1); второй — с выбором преобразования Φ в том же выражении (1), по которым комбинируются контурные модели, полученные на основе предварительных преобразований, и третий — с выбором фазовых координат для фазовых изображений. Отметим сразу, что пока еще не существует регулярных методов решения перечисленных вопросов. Однако сочетание некоторой научной интуиции, основанной на опыте работ по распознаванию речевых сигналов, с известными положениями в психоакустике и в электрофизиологии, а также с некоторыми аналогиями, заимствованными из дифференциальной геометрии и механики,

позволяет построить, по мнению авторов, весьма перспективные с точки зрения распознавания структурные модели речевых сигналов.

2. Предварительные преобразования речевого сигнала

Как было отмечено, целью предварительной обработки сигналов является создание контурных моделей, из которых потом строятся структурные модели согласно определениям, приведенным выше.

Отметим сразу, что арсенал способов обработки довольно широк и включает как линейные, так и нелинейные функциональные трансформации речевых сигналов. В принципе, здесь допустимы любые функциональные преобразования, имеющие смысл с точки зрения распознавания. В соответствующей литературе можно найти множество примеров применения преобразований речевых сигналов, которые (по нашей терминологии) приводят к контурным моделям. Ниже мы выделим некоторые из них, как наиболее отвечающие поставленной задаче, а также представлениям о свойствах самих речевых сигналов и о механизмах их слуховой обработки.

Условно все преобразования можно делить на две основные группы: преобразования, основанные на спектральном подходе, и преобразования, основанные на временном подходе. И хотя спектральный подход лучше соответствует данным о механизме работы слухового анализатора, нельзя не учесть того факта, что изучение временной картины сигнала и особенно «тонкой» временной его структуры позволяет добиться хороших результатов как при анализе, так и при распознавании [26, 27]. Кроме того, временной подход в ряде случаев допускает уменьшение аппаратных затрат.

Среди преобразований, основанных на спектральном подходе, выделим формирование огибающих $A_n(t)$ в частотных полосах сигнала $s(t)$ при помощи детекторов и фильтров низких частот. Рациональной верхней цифрой количества используемых полосовых фильтров надо считать $n=24$; при этом фильтры берутся с полосами пропускания, близкими к критическим полосам слуха. Кроме того, выделяются моментные функции $M_k(t)$, позволяющие рационально объединять в одну модель огибающие $A_n(t)$, где

$$M_k(t) = \sum_n \omega_n^k A_n(t), \quad k = 0, 1, 2, \dots$$

При меньшем количестве используемых фильтров рационально привлечь преобразования для выделения функций мгновенных частот $\omega_n(t)$ в частотных полосах. Математически эти функции определяются так:

$$\omega_n(t) = \varphi'_n(t) = \frac{d}{dt} \left[\operatorname{arctg} \frac{\hat{s}_n(t)}{s_n(t)} \right],$$

где $s_n(t)$ — сигнал на выходе n -го фильтра, а $\hat{s}_n(t)$ — преобразование Гильберта для этого сигнала. Аппаратурно они выделяются

при помощи ромеров или частотных дискриминаторов. Функции $\omega_n(t)$ отражают в какой-то мере фазовые соотношения в сигнале, и с чисто формальной точки зрения только совокупность функций $A_n(t)$ и $\omega_n(t)$ дает полное описание сигнала—описание без потерь информации. Хотя фазовые соотношения в речевых сигналах как бы не играют особенной роли при их восприятии, работы по исследованию так называемых фазовых вокодеров [21] указывают на целесообразность использования функций $\omega_n(t)$. Эти функции в ряде случаев несут даже больше информации о сигнале, чем огибающие $A_n(t)$.

На основе временного подхода считаем рациональным использование самого сигнала $s(t)$ и выходных сигналов полосовых фильтров, т. е. $s_n(t)$.

Кроме того, к выделяемым преобразованиям следует отнести дифференцирование перечисленных функций до второго порядка включительно. Из-за затруднений с аппаратурной реализацией дифференциаторов (с ростом порядка дифференцирования сильно возрастает влияние шумов) вряд ли целесообразно выделять производные более высоких порядков. Исключения, конечно, составляют теоретические построения.

В заключение приведем список нами предпочитаемых контурных моделей речевого сигнала: огибающие $A_n(t)$, мгновенные частоты $\omega_n(t)$, моментные функции $M_n(t)$, сам сигнал $s(t)$, сигналы в частотных полосах $s_n(t)$ и их первые и вторые производные. Вместе с тем легко записать в явном виде выражения соответствующих этим моделям трансформаций L_1, L_2, \dots . Однако в этом нет необходимости, так как эта запись имела бы лишь иллюстративное значение.

3. Некоторые структурные связи

Как уже указывалось, ни определение структурной связи с явным временем, ни определение структурной связи с неявным временем не позволяют однозначно выбрать правило объединения контурных моделей. Исследователю предоставляется некоторая свобода действий, которой, естественно, надо пользоваться осторожно и разумно. Ниже введем и рассмотрим некоторые возможные структурные связи, на наш взгляд, наиболее интересные и наиболее перспективные как с точки зрения их информационных свойств и возможностей аппаратурного выделения, так и с точки зрения современных представлений о моделях механизмов слуха. При этом определим и правила объединения контурных моделей.

В дальнейшем будем предполагать, что контурные модели, из которых «конструируется» структурная модель, известны на основе выбранных предварительных преобразований сигнала $s(t)$. Пусть тогда $L_1[s(t)] = x_1(t) = x$, $L_2[s(t)] = x_2(t) = y$, $L_3[s(t)] = x_3(t) = z$, ... Ограничимся рассмотрением трех контурных моделей $x_1(t)$, $x_2(t)$, $x_3(t)$, связанных между собой соотношениями: $x_2(t) = \dot{x}_2(t)$, $x_3(t) = \dot{x}_2(t) = \dot{x}_1(t)$, или в более компактной записи $y = \dot{x}$, $z = \dot{y} = \ddot{x}$. Тогда, если x в механической интерпретации — координата, соответствующая положению материальной точки, то y — скорость движения, а z — ускорение. Эти три параметра

достаточно полно характеризуют само движение. Между x , y и z образуем следующие структурные связи с явным временем:

$$\left. \begin{aligned} \delta &= \frac{y}{x} = \frac{\dot{x}}{x}, \\ \kappa &= \frac{z}{x} = \frac{y}{x} \frac{dy}{dx} = \frac{\ddot{x}}{x}, \\ \rho &= \frac{z}{y} = \frac{dy}{dx} = \frac{\ddot{x}}{\dot{x}} = \frac{\kappa}{\delta}, \\ \lambda &= \frac{z}{y} \frac{x}{y} = \frac{x}{y} \frac{dy}{dx} = \frac{\ddot{x}x}{\dot{x}^2} = \frac{\kappa}{\delta^2} = \frac{\rho}{\delta}. \end{aligned} \right\} (2)$$

(Для этих конкретных структурных связей мы ввели новые обозначения δ , κ , ρ , λ , хотя в общем случае структурные связи с явным временем обозначили через $\xi(t)$).

Приведем определения введенных структурных связей [14] и дадим некоторое истолкование их физического смысла.

1. Диссипант δ связан с представлениями о диссипации энергии: при $\delta > 0$ энергия возрастает, а при $\delta < 0$ — убывает, т. е. имеет место диссипация энергии. Если воспользоваться приближенным равенством $\delta \cong \frac{\Delta x/x}{\Delta t}$, то можно дать следующее определение:

диссипант представляет собой скорость относительного изменения процесса. Для функции пропорционального роста (экспоненциальные процессы) диссипант равен постоянному числу, т. е. инвариантен ко времени, а для постоянной он равен нулю.

2. Консервант κ связан с представлениями о «консервативности энергии»: если $\kappa = \text{const} < 0$, энергия «консервативна», т. е. в среднем во времени не изменяется, а непрерывно преобразовывается из потенциальной в кинетическую и обратно. Перепишем выражение для κ в виде $\kappa = \frac{ydy}{xdx} = \frac{dy^2}{dx^2}$, легко заключить, что консервант представляет собой производную одного вида энергии по другому виду. Приближенное равенство $\kappa \approx \frac{\Delta x_0/x}{\Delta t}$

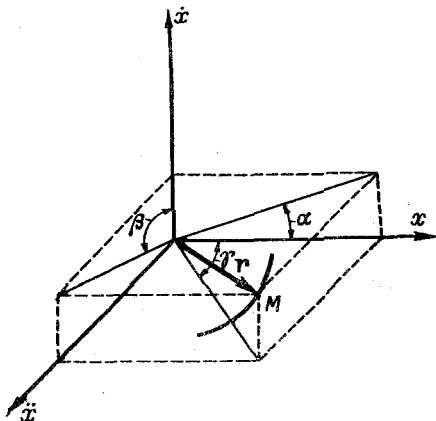
указывает на то, что консервант имеет также смысл ускорения относительного изменения процесса. Для гармонического колебания с частотой ω консервант $\kappa = -\omega^2$, а для экспоненциального процесса с декрементом (инкрементом) α консервант $\kappa = \alpha^2$, т. е. консервант инвариантен к амплитуде и фазе гармонического колебания, к знаку показателя экспоненциального процесса и ко времени для обоих процессов.

3. Кондисант ρ представляет собой отношение консерванта к диссипанту, именно этим объясняется данное словообразование. С другой стороны, как вытекает из (2), кондисант равен производной скорости по координате. Приближенное равенство $\rho \approx$

$\approx \frac{\Delta y/y}{\Delta t}$ позволяет трактовать кондисант как скорость относительного изменения скорости.

4. Кондисквадрат дает отношение консерванта к диссипанту в квадрате, это обстоятельство и учитывается в названии. Из приближенного равенства $\lambda \approx \frac{\Delta y/y}{\Delta x/x}$ следует: кондисквадрат дает отношение относительного изменения скорости процесса к относительному изменению самого процесса. Для экспоненциального процесса кондисквадрат равен единице при любом показателе, будь то декремент или инкремент.

Рис. 1. Геометрическая интерпретация некоторых структурных связей



Дадим еще одну, геометрическую интерпретацию введенных структурных связей. Для этого изобразим движение материальной точки M по соответствующей траектории в трехмерном фазовом пространстве x, \dot{x}, \ddot{x} (рис. 1).

Движение точки M по траектории полностью определяется заданием радиус-вектора r , которое равносильно заданию модуля вектора r и трех углов α, β, γ (рис. 1) между соответствующими координатными осями и проекциями вектора r на координатные плоскости. Очевидно:

$$\operatorname{tg} \alpha = \frac{\dot{x}}{x} = \delta, \quad \operatorname{tg} \beta = \frac{\ddot{x}}{\dot{x}} = \rho, \quad \operatorname{tg} \gamma = \frac{\ddot{x}}{x} = \kappa.$$

Перечисленные структурные связи могут использоваться как связи с явным временем, т. е. как новые процессы $\delta(t), \kappa(t), \rho(t), \lambda(t)$, а также как связи с неявным временем, т. е. как фазовые изображения видов $x, \delta; x, \kappa; x, \rho; x, \lambda$ и более сложных видов $\kappa, \delta; \rho, \delta$ и т. д. К числу применяемых фазовых изображений следует отнести также и элементарное фазовое изображение x, \dot{x} .

Может оказаться целесообразным также привлечение образцовых структурных связей, т. е. связей вида:

$$\frac{1}{\delta} = \frac{x}{\dot{x}}, \quad \frac{1}{\kappa} = \frac{x}{\ddot{x}}, \quad \frac{1}{\rho} = \frac{\dot{x}}{\ddot{x}}, \quad \frac{1}{\lambda} = \frac{\dot{x}^2}{\ddot{x}\dot{x}}.$$

Однако опять возникает естественный вопрос: как по свойствам рассматриваемого процесса выбрать наиболее подходящую структурную связь, хотя бы из числа перечисленных. Исчерпывающего ответа на этот вопрос пока нет. Можем указать лишь на один возможный путь дальнейших исследований. Рассматривая изучаемый процесс как случайный, можно построить его закон распределения, а также законы распределения структурных связей — диссипанта, консерванта и т. д. Можно отдать предпочтение тому виду связи, который приводит к уменьшению энтропии распределения (по сравнению с энтропией исходного распределения). То, что действительно это имеет место, можно показать даже аналитически на примере нормального распределения стационарного процесса [15]. Практическое решение такой задачи очевидно потребует громоздких вычислений, но несомненно является доступным.

4. Соображения об адекватности структурной модели речевого сигнала моделям механизмов слуха

Постановка вопроса о целесообразности привлечения новой математической модели сигнала несомненно правомерна, и на этот вопрос должен быть дан ответ. Ниже и будут приведены соответствующие разъяснения и соображения.

Нам предстоит сопоставить контурные модели со структурными; по существу нужно ответить на вопрос: почему объединение, скажем, двух фазовых координат в одну структурную связь лучше, чем две отдельные фазовые координаты.

Для ясности можно этот вопрос упростить еще больше. Пусть мы располагаем значениями координаты и скорости движущейся материальной точки. Нужно ли конструировать некоторую структурную связь между ними и рассматривать именно эту связь, вместо того чтобы рассматривать путь и скорость раздельно?

На этот последний вопрос ответ представляется ясным. Дело в том, что совокупности значений координаты и скорости являются описаниями конкретного процесса (конкретной реализации, конкретного образа); для описания же ансамбля конкретных реализаций по их характерным свойствам нужно воспользоваться описанием связей между координатой и скоростью (или другими фазовыми координатами). Введение структурной связи позволяет, не отображая каждой конкретной реализации, описать их ансамбль.

Структурная модель обобщает контурную, и последняя представляет собой наипростейший вариант первой; каждое обобщение таит в себе новые возможности и новые преимущества, которые часто обнаруживаются не сразу.

Использование даже простейших структурных моделей (например, фазовое изображение x, \dot{x}) приводит к новому математическому аппарату синтеза алгоритмов обработки сигнала и, следовательно, к новым типам устройств. Здесь уместна аналогия со взаимосвязью между временной картиной сигнала и его спектром. Хорошо известно, что вся информация о сигнале содержится в его временной картине, однако использование спектра позволяет относительно легко извлекать такую информацию, которую трудно получить на основе непосредственного изучения временной картины.

Можно высказать предположение, что человеческие органы восприятия используют также некоторые структурные связи. Хотя более или менее полной теории восприятия речи пока еще нет, нет еще также математических формулировок задачи исследования и в соответствии с этим не установлены экспериментальные методы, мы сделаем попытку перечислить те математические операции, которым предположительно подвергаются сигналы в слуховом анализаторе. Этот материал полностью основан на известных по литературным данным соображениях из областей, охватывающих психологию, психоакустику и физиологию.

Частотная фильтрация речевого сигнала и последующее ее выделение низкочастотных огибающих являются общепризнанными операциями, осуществляемыми слуховым анализатором. Кроме того, известен закон Вебера—Фехнера, в соответствии с которым сила ощущения пропорциональна логарифму интенсивности раздражения. Электрофизиологические данные указывают также, что ответная реакция нейрона на периферии слуховой системы пропорциональна логарифму средней за некоторый промежуток времени энергии стимула [18]. Вполне вероятно, здесь мы имеем дело с преобразованиями вида $\ln x$, где x — низкочастотная огибающая сигнала в частотном канале.

Нейрофизиологические и психоакустические данные относительно обработки огибающей стимула показывают также [11], что человек обнаруживает как положительные, так и отрицательные приращения в огибающей стимула, причем переменная, на которую он реагирует, близка к скорости изменения огибающей. Это позволяет предполагать, что преобразование сигнала соответствует дифференцированию огибающей стимула.

На основании вышеизложенного можно считать, что в слуховом анализаторе осуществляются по крайней мере следующие математические действия над сигналами: дифференцирование, квадрирование, логарифмирование и, естественно, суммирование. При помощи этих математических действий можно определить введенные выше структурные связи (с точностью до постоянного множителя c), например:

$$\left. \begin{aligned} (\ln x^2)' &= cd, & (\ln \dot{x}^2)' &= cp, \\ \ln cx &= \ln [(\ln x^2)'] + \ln [(\ln \dot{x}^2)'], \\ \ln \lambda &= \ln [(\ln \dot{x}^2)'] - \ln [(\ln x^2)']. \end{aligned} \right\} \quad (3)$$

Весьма вероятно, что слуховой анализатор реализует указанные (или им подобные) преобразования сигналов; высказанное предположение доступно экспериментальной проверке.

Отметим, что выражения (3) могут служить в качестве алгоритмов при синтезе схем устройств для выделения соответствующих функций, хотя в настоящее время возможно и прямое моделирование соотношений (2).

Ниже приведем несколько примеров, основанных на литературных источниках, показывающих успешное использование некоторых моделей (сводимых к структурным) речевых сигналов для их распознавания или сегментации.

Рассмотрим с точки зрения понятий о структурных связях хорошо известную фонетическую функцию Пирогова [16], предложенную на основании прямых психоакустических экспериментов и литературных данных о работе слухового анализатора в качестве модели речевых сигналов для их распознавания. Согласно

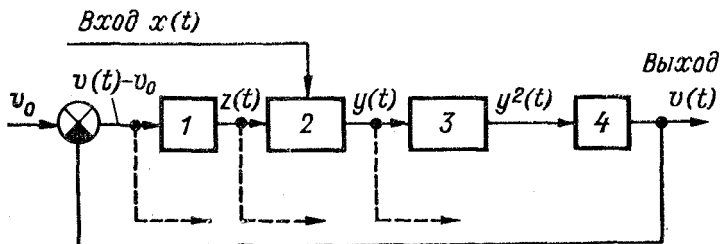


Рис. 2. Структурная схема модели адаптивного механизма
1 — интегратор, 2 — регулируемый усилитель, 3 — квадратор, 4 — фильтр низких частот

[16] фонетическая функция $P(\omega, t)$ связана со спектральной функцией $S(\omega, t)$ речевого сигнала следующим соотношением:

$$P(\omega, t) = \ln \frac{S(\omega, t)}{S(\omega, t - \tau)}$$

Последнее выражение можно преобразовать к виду

$$P(\omega, t) = \ln S(\omega, t) - \ln S(\omega, t - \tau) = \int_{t-\tau}^t \frac{\dot{S}(\omega, u)}{S(\omega, u)} du.$$

Если воспользоваться определением структурной связи [диссипанта (3)], то получим:

$$P(\omega, t) = \int_{t-\tau}^t \delta_s(\omega, u) du.$$

Другими словами, фонетическая функция $P(\omega, t)$ является интегральной оценкой диссипанта спектральной функции. Имея в виду, что при аппаратном выделении в качестве спектральной функции используется совокупность низкочастотных огибающих в частотных каналах, здесь уместно говорить о диссипанте огибающей.

По существу и в более ранней работе [7] предлагается использовать для описания речевых сигналов при их распознавании диссипант огибающей в частотном канале и его интегральную оценку, т. е. $\delta(t) = \frac{d}{dt} [\ln x(t)]$ и $J = \int_{t_1}^{t_2} \delta(t) dt$. Та же самая интегральная оценка диссипанта огибающей в частотном канале используется для сегментации речевых сигналов на фонемы в работе [19].

Результаты исследования адаптивных механизмов слухового анализатора позволяют построить простейшую модель такого рода механизмов [17], подобную хорошо известной системе автоматического регулирования усиления (рис. 2). Такие устройства могут быть использованы для членения речевых сигналов на динамические и квазистационарные участки, т. е. для их предварительной сегментации [17]. Нетрудно показать, что преобразования сигналов в устройстве аналогичны преобразованиям, приводящим к структурным моделям.

Уравнения, связывающие выходные и входные процессы в отдельных блоках устройства, имеют вид [17]:

$$\left\{ \begin{array}{l} y(t) = K(z) x(t), \quad K_1^2(z) = 10^{-\alpha z}, \\ z(t) = \frac{1}{T} \int_{-\infty}^t \frac{v(\xi) - v_0}{v_0} d\xi, \\ v(t) = \frac{\beta}{\tau} \int_{-\infty}^t y^2(\eta) \exp\left(\frac{\eta - t}{\tau}\right) d\eta, \end{array} \right.$$

где K — коэффициент передачи регулируемого усилителя, α , β , τ , v_0 , T — постоянные, определяющие характеристики соответствующих блоков.

Система уравнений (4) сводится к нелинейному дифференциальному уравнению второго порядка относительно переменной $z(t)$ [17]

$$\ddot{z} + \frac{1}{\tau} \dot{z} = \frac{\beta 10^{-2\alpha z}}{\tau T v_0} x^2(t) - \frac{1}{\tau T},$$

аналитическое решение которого в общем виде затруднено (оно допускает численное интегрирование на ЭВМ). Однако некоторые выводы относительно зависимости z от x можно сделать, решив уравнение при некоторых упрощающих допущениях. Допустим, что входное воздействие $x(t)$ узкополосное. Между прочим, такое допущение почти не ограничивает общности задачи, так как сигналы на систему, как правило, подаются с выходов довольно узкополосных полосовых фильтров [17]. Допустим также, что 4 на рис. 2 — идеальный фильтр нижних частот, а его полоса

пропускания охватывает весь частотный спектр огибающей воздействия $x(t)$. Тогда на основании этих допущений следует:

$$x(t) = X(t) \cos \varphi(t), \quad y(t) = Y(t) \cos \varphi(t), \quad v(t) = Y^2(t),$$

где $X(t)$ и $Y(t)$ — огибающие функции процессов $x(t)$ и $y(t)$ соответственно, а $\varphi(t)$ — их фазовая функция.

Система уравнений (4) приобретает вид

$$\begin{cases} Y(t) = K(z) X(t), & K(z) = 10^{-\alpha z}, \\ z(t) = \frac{1}{T} \int_{-\infty}^t \frac{v(\xi) - v_0}{v_0} d\xi, \\ v(t) = Y^2(t) \end{cases}$$

и сводится к уравнению Бернулли

$$\dot{Y} + \frac{\alpha}{Tv_0} Y^3 - \left(\frac{\alpha}{T} + \frac{\dot{x}}{x} \right) Y = 0.$$

Решая это уравнение классическими методами получаем:

$$Y^2(t) = \frac{Tv_0 X^2(t) \exp\left(\frac{2\alpha}{T} t\right)}{2\alpha \int X^2(t) \exp\left(\frac{2\alpha}{T} t\right) dt}.$$

Нетрудно видеть, что данное решение соответствует определению введенного диссипанта функции $\int X^2(t) \exp\left(\frac{2\alpha}{T} t\right) dt$.

Следует отметить, что приведенные примеры и соображения отражают новейшее направление в построении алгоритмов и технических средств обработки речевых сигналов с целью их распознавания и сегментация и являются хорошо обоснованными с точки зрения психоакустики и электрофизиологии. Нами же предлагаемые структурные модели являются обобщением частных примеров приведенных выше.

Таким образом, сказанное выше дает достаточно весомые основания для заключения о предпочтительности структурной модели сигнала перед контурной в деле распознавания речевых сигналов.

Глава вторая

ПРИЗНАКИ ДЛЯ РАСПОЗНАВАНИЯ

5. Общие положения

Процесс распознавания речевых сигналов невозможен без предшествующего процесса измерения их параметров. При выборе информационных параметров необходимо считаться с двумя противоречивыми требованиями. С одной стороны, для осуществ-

ления надежного распознавания необходимо сохранить исходную информацию. С другой стороны, с точки зрения простоты технической реализуемости, а также по экономическим соображениям количество измеряемых параметров и точность их измерения должны быть по возможности небольшими. В связи с этим исследователи стремятся найти компактные описания речевых сигналов, удовлетворяющие поставленным требованиям. Общее решение этого вопроса представляет значительные трудности. К настоящему времени еще нет законченной математически строгой и практически реализуемой методики выбора систем эффективных признаков. Поэтому сейчас основная роль в разработке компактных и эффективных описаний речевых сигналов принадлежит экспериментально-эвристическим поискам. Наиболее целесообразными, естественно, надо считать те методы описания речевых сигналов, которые опираются на современные представления о характере переработки информации в слуховом анализаторе человека. Слуховой анализатор человека при распознавании речи осуществляет измерение параметров инвариантно к разным характеристикам самого сигнала: к тембру, темпу, интенсивности речи и т. д. Именно эти особенности мы считаем основными при выборе системы признаков.

Информативный признак, прежде всего, должен удовлетворять двум основным требованиям. Во-первых, он должен сохранять неизменным свое значение для всех сигналов, принадлежащих данному классу, т. е. быть инвариантным к преобразованиям подмножества параметров, порождающих данный класс, а во вторых, изменять свое значение при переходе к сигналам из других классов, т. е. зависеть от другого подмножества параметров. В этом смысле информативный признак должен являться нетривиальным инвариантом группы преобразований, порождающей данный класс сигналов.

Задача нахождения информативного признака адекватна задаче поиска структурной модели сигнала, инвариантной к параметрам, порождающим рассматриваемый класс. Однако в случае речевого сигнала эти параметры не постоянны на интервале времени, равном длительности сигнала, за исключением может быть квазистационарных участков (но и то только условно). Кроме того, априорная информация о свойствах этих параметров чрезвычайно бедна. В этих условиях целесообразно нахождение информативных признаков вести путем поиска числовых характеристик структурной модели сигнала в виде некоторых ее функционалов.

В связи с тем что введены два вида структурных связей: с явным временем и с неявным временем — числовые их характеристики удобно искать для каждого вида в отдельности. В последующих параграфах настоящей главы и рассматриваются основы определения информативных признаков — числовых характеристик обоих видов структурных связей. Рассматриваются следующие системы признаков: а) квазитопологические характеристики фазовых изображений, т. е. числа пересечений фазовыми изображениями некоторых граничных линий на фазовой плоскости; б) время пребывания изображающей точки фазового изображения в секторах фазовой плоскости; в) признаки, основанные на кодировании секторов фазовой плоскости фазовых изо-

бражений; г) интегральные оценки структурных связей с явным временем.

Измерение информативных параметров речевых сигналов тесно связано с вопросами определения длительностей минимальных отрезков сигналов, для которых целесообразно проводить само измерение. Однако членение сигнала на отрезки согласно принятому критерию качества относится уже к проблеме сегментации. В связи с этим в конце главы рассматриваются некоторые вопросы сегментации речевых сигналов на основании их структурных свойств.

6. Топологические инварианты

Каждому определенному сигналу соответствует на фазовой плоскости определенная фазовая траектория, параметры которой жестко связаны с параметрами сигнала. Изменение любого из

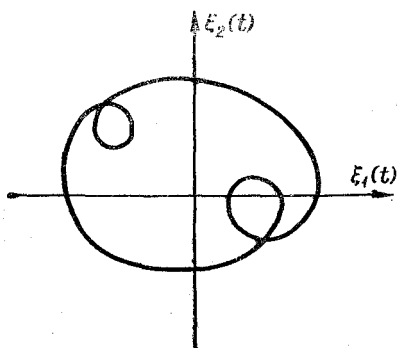


Рис. 3. Фазовое изображение

параметров сигнала (это изменение может быть вызвано модуляцией параметров полезного сигнала, наличием помех и т. п.) приведет к изменению изображающей траектории, которое можно рассматривать в качестве искажения исходной траектории.

Задача заключается в нахождении такого алгоритма обработки сигнала, который позволил бы отыскать некоторое соответствующее известному варианту сигнала число, не зависящее от ряда параметров сигнала. Тем самым постулируется возможность отыс-

кания такого числа (меры на фазовой плоскости), которое соответствовало бы некоторому ансамблю сигналов, каждая из реализаций которого имеет специфическую, присущую только ей фазовую траекторию.

Введем упрощающее допущение — рассматриваемые фазовые траектории сигналов замкнуты (в дальнейшем указанное допущение будет снято); кроме того, выберем на фазовой плоскости некоторое начало отсчета, относительно которого будем рассматривать поведение фазовых траекторий.

Точку начала отсчета назовем полюсом. Выбор оптимальных координат полюса представляет собой самостоятельную задачу; пока, не теряя общности рассмотрения, поместим полюс в начало координат.

Если для каждой из реализаций фазовых траекторий координаты полюса считать неизменными, то различия между ними будут выражены только в особенностях формы траектории. Так как в каждый момент времени эти различия могут существовать, то общее между фазовыми траекториями отдельных реализаций

следует искать в классе интегральных инвариантов, учитывающих поведение траектории на всем временном интервале задания сигнала.

Наибольшими искажениями, которым может подвергнуться фазовая траектория сигнала, являются непрерывные ее деформации. Деформация называется непрерывной, если траектории любых из реализаций сигнала могут быть непрерывно и однозначно совмещены друг с другом, минуя полюс. Пересечение с полюсом означает разрыв траектории. Свойства траекторий, остающиеся неизменными при таких преобразованиях, принято относить к топологическим инвариантам. Тем самым топологические инварианты характеризуют траектории (и соответствующие им сигналы) с наиболее важной стороны. Их наличие связано с тем, что, несмотря на искажения сигнала, приводящие к пересечению подмножеств сигналов разных классов, всегда остается некоторый «минимум организованности», связанный с априорными сведениями о структуре сигнала. В один и тот же класс попадают только те траектории (сигналы), которые характеризуются одним и тем же значением инварианта (числовой характеристики). В общем случае числовая характеристика фазовой траектории $X = \xi_1(t)$, $Y = \xi_2(t)$ или, короче, (ξ_1, ξ_2) (рис. 3) может быть представлена в виде криволинейного интеграла [22]:

$$J = \oint_G P(\xi_1, \xi_2) d\xi_1 + Q(\xi_1, \xi_2) d\xi_2, \quad (5)$$

где P, Q — некоторые функции фазовых координат, G — контур, вдоль которого ведется интегрирование. (В частном случае под G может пониматься фазовая траектория.)

С учетом приведенных рассуждений о деформации фазовых траекторий определение интегрального инварианта, описывающего совокупность реализаций, сводится к отысканию таких функций P и Q , для которых криволинейный интеграл (5) не зависит бы от пути интегрирования.

Условие независимости интеграла от пути интегрирования $J = \text{invar} \{G\}$ заключается в существовании функции $U(\xi_1, \xi_2)$, такой, что

$$dU(\xi_1, \xi_2) = Pd\xi_1 + Qd\xi_2.$$

Необходимым и достаточным условием существования функции U является выполнение равенства $\frac{\partial P}{\partial \xi_2} = \frac{\partial Q}{\partial \xi_1}$ при условии непрерывности частных производных [22].

После того как определено положение полюса и установлен вид функций P и Q , можно утверждать, что все сигналы подразделяются на классы в зависимости от значений J . Для всех реализаций одного класса значение интегральной меры остается неизменным.

Так как мы рассматриваем только непрерывные деформации замкнутых траекторий, то значение J определяется особенностями фазовой траектории относительно полюса.

В качестве возможной интегральной меры, удовлетворяющей всем предъявленным выше требованиям, может быть использовано понятие порядка кривой (фазовой траектории) отно-

сительно точки (полюса) (т. е. вращение радиального поля) или понятие индекса векторного поля относительно замкнутой кривой [2, 24] (вращение касательного векторного поля).

Порядок кривой определяется углом поворота против часовой стрелки радиус-вектора изображающей точки относительно полюса за время наблюдения сигнала и выражается в относительных единицах.

Если фазовая траектория задана параметрически через $(\xi_1(t))$ и $\xi_2(t)$, то этот угол (обозначен через j_r) определится соотношением:

$$j_r = \frac{1}{2\pi} \oint_G d \left[\arctg \frac{\xi_1}{\xi_2} \right] = \frac{1}{2\pi} \int_{t_0}^{t_0+T} \frac{\left| \begin{array}{cc} \dot{\xi}_1 & \xi_2 \\ \xi_1 & \dot{\xi}_2 \end{array} \right|}{\xi_1^2 + \xi_2^2} dt, \quad (6)$$

где T — продолжительность наблюдения сигнала.

Так как фазовая траектория замкнута, то j_r есть целое число.

Параметр j_c — вращение касательного векторного поля вдоль замкнутой фазовой траектории — определяется по известной формуле индекса Пуанкаре [2]:

$$j_c = \frac{1}{2\pi} \oint_G d \left[\arctg \frac{\dot{\xi}_1}{\dot{\xi}_2} \right] = \frac{1}{2\pi} \int_{t_0}^{t_0+T} \frac{\left| \begin{array}{cc} \dot{\xi}_1 & \dot{\xi}_2 \\ \xi_1 & \xi_2 \end{array} \right|}{\dot{\xi}_1^2 + \dot{\xi}_2^2} dt. \quad (7)$$

Очевидно, что $j_c = j_r$, если j_c вычисляется для фазового изображения ξ_1, ξ_2 , а j_r — для фазового изображения $\dot{\xi}_1, \dot{\xi}_2$. Тем самым вычисление инвариантов (6), (7) может быть сведено к вычислению только порядков кривых на различных фазовых плоскостях.

Укажем еще один вид алгоритма (7), который позволяет более полно вскрыть его внутреннее содержание.

Введем в рассмотрение кривизну K фазовой траектории:

$$K = \frac{\left| \begin{array}{cc} \dot{\xi}_1 & \dot{\xi}_2 \\ \xi_1 & \xi_2 \end{array} \right|}{(\xi_1^2 + \xi_2^2)^{3/2}}$$

и скорость движения изображающей точки v :

$$v = \sqrt{\dot{\xi}_1^2 + \dot{\xi}_2^2}.$$

Тогда

$$j_c = \frac{1}{2\pi} \int_{t_0}^{t_0+T} K(t) v(t) dt = \frac{1}{2\pi} \oint_G K dl,$$

где l — длина траектории.

Таким образом, принятая интегральная мера выражает величину площади, ограниченной кривизной фазовой траектории, когда в качестве аргумента выступает длина траектории.

В качестве иллюстрации на рис. 4 представлены фазовые изображения ряда сигналов, которые по признаку тождественности j_r могут быть отнесены к одному и тому же классу.

До сих пор рассмотрение ограничивалось только замкнутыми фазовыми траекториями, для которых и были определены интегральные инварианты—топологические инварианты (топологические характеристики). Вместе с тем в большинстве случаев требование замкнутости траекторий не соблюдается на интервале наблюдения сигнала. Это, в свою очередь, влечет за собой приближенный характер полученных строгих соотношений. Однако при достаточно большом времени анализа, когда $j_r \gg 1$, $j_e \gg 1$, влияние незамкнутости траекторий на результаты вычислений нивелируется.

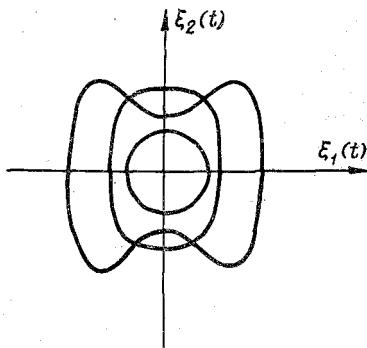


Рис. 4. Фазовые изображения, тождественные по признаку j_r

7. Квазитопологические характеристики

Определение указанных в § 6 интегральных инвариантов связано с большими затруднениями, так как для их вычисления необходимо выполнение таких аналоговых операций, как многократное дифференцирование, перемножение и деление функций. Можно обойти указанные затруднения,

если обратиться к квазитопологическим характеристикам, из которых наибольший интерес представляют числа пересечений фазовой траекторией некоторой кривой (линии связи), заданной уравнением $F(\xi_1, \xi_2) = 0$ (рис. 5). Подсчет числа

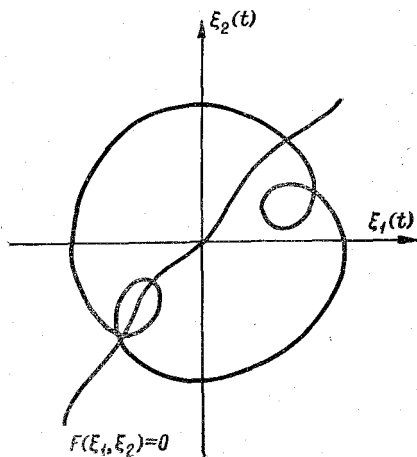


Рис. 5. Пересечения фазовым изображением линии связи

пересечений может быть осуществлен путем подсчета числа переходов через нуль функции

$$z(t) = F[\xi_1(t), \xi_2(t)]. \quad (8)$$

Утверждается, что всегда можно подобрать такие линии связи, число пересечений которых инвариантно по отношению к непрерывным и однозначным деформациям фазовой траектории, и в этом смысле числа пересечений тождественны ранее рассмотренным интегральным оценкам (топологическим инвариантам).

Легко видеть из простых геометрических представлений, что в случае фазового изображения вида $X = \xi_1(t) = \xi(t)$, $Y = \xi_2(t) = \xi(t)$ число пересечений оси абсцисс ($z(t) = \xi(t)$) совпадает

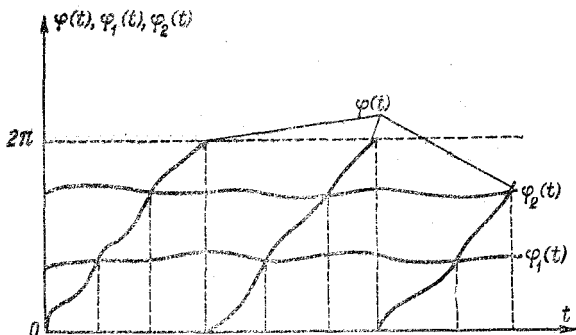


Рис. 6. Интерпретация пересечений фазовым изображением линии связи

с удвоенным вращением касательного поля, а число пересечений оси ординат ($z(t) = \xi(t)$) — с удвоенным вращением радиального поля (относительно начала координат).

В случае же любых фазовых изображений такой выбор линий связи составляет самостоятельную и довольно трудно разрешимую задачу. Однако в нем нет особой необходимости. Можно использовать в принципе любые линии связи, в том числе прямые. При произвольно выбранных линиях связи следует уже говорить о квазитопологических инвариантах, введение которых позволяет расширить множество инвариантных признаков за счет свободного выбора самих линий связи.

Пересечения фазовым изображением линий связи можно трактовать также, как пересечения функции $\varphi(t)$ кривых $\varphi_1(t)$ и $\varphi_2(t)$ (см. рис. 6). Функция $\varphi(t)$ соответствует моментному значению угла поворота в положительном направлении радиус-вектора изображающей точки относительно полюса, т. е.

$$\varphi(t) = \arctg \frac{\xi_2(t)}{\xi_1(t)}. \quad (9)$$

Кривые $\varphi_1(t)$ и $\varphi_2(t)$ на рис. 6 соответствуют одной линии связи на фазовой плоскости (при одном полном обороте радиус-вектора фазовое изображение по крайней мере дважды пересекает линию связи).

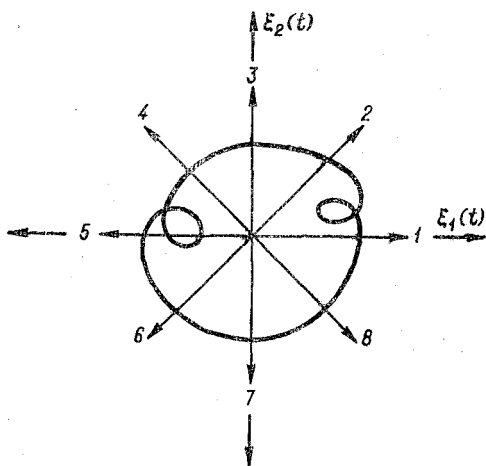


Рис. 7. Пересечения фазовым изображением лучей связи

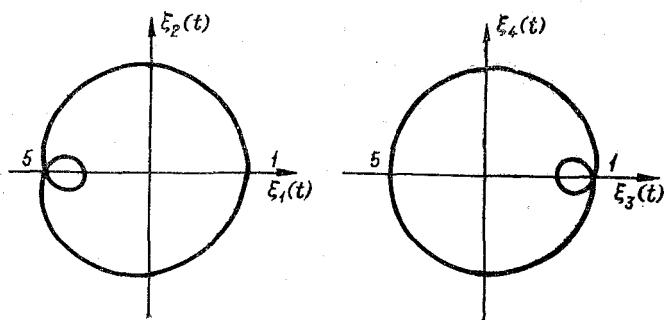


Рис. 8. Иллюстрация целесообразности перехода к лучам связи

Целесообразно использовать лучи в качестве линий связи, как это показано на рис. 7, где вся фазовая плоскость расчленена на ряд областей восемью лучами. При этом расширение множества инвариантных признаков достигается путем изменения тангенсов углов наклона лучей. Кроме того, целесообразность

перехода к лучам связи легко демонстрируется на примере рассмотрения фазовых изображений ξ_1, ξ_2 и ξ_3, ξ_4 на рис. 8. Для обоих изображений вращения радиального и касательного полей совпадают, совпадают также и числа пересечений осей абсцисс и ординат. В то же время существенная разница между изображениями легко раскрывается, если отдельно рассмотреть число пересечений лучей 1 и 5.

Таким образом, переход от линий к лучам связи позволяет вскрыть дополнительную информацию о структуре сигнала, представленного на одной фазовой плоскости. Следует подчеркнуть, что при этом также сужается круг допустимых преобразований, относительно которых достигается инвариантность. Введение лучей, в частности, исключает инвариантность по отношению к фазовым составляющим сигнала (пример на рис. 8).

Очевидно, кривые $\varphi_1(t), \varphi_2(t)$ (см. рис. 6) в случае использования лучей связи становятся прямыми, параллельными оси времени. В то же время функция $z(t)$ в выражении (8), для которой подсчет числа переходов через нуль может быть использован вместо подсчета пересечений, приобретает вид

$$z(t) = \xi_2(t) \pm k\xi_1(t) \quad (10)$$

при условиях, определяющих сектор плоскости, в котором находится рассматриваемый луч. (Здесь k — тангенс угла наклона луча). Так, например, в случае восьми лучей связи (см. рис. 7) имеем:

для первого луча	$z(t) = \xi_2(t), \xi_1(t) > 0;$	}	(11)
для второго луча	$z(t) = \xi_2(t) - k\xi_1(t), \xi_1(t) > 0;$		
для третьего луча	$z(t) = \xi_1(t), \xi_2(t) > 0;$		
для четвертого луча	$z(t) = \xi_2(t) + k\xi_1(t), \xi_1(t) < 0;$		
для пятого луча	$z(t) = \xi_2(t), \xi_1(t) < 0;$		
для шестого луча	$z(t) = \xi_2(t) - k\xi_1(t), \xi_1(t) < 0;$		
для седьмого луча	$z(t) = \xi_1(t), \xi_2(t) < 0;$		
для восьмого луча	$z(t) = \xi_2(t) + k\xi_1(t), \xi_1(t) > 0.$		

Выбор количества лучей связи и тангенсов их углов наклона — вопрос пока не решенный в общем виде. Можно высказать следующие соображения относительно его решения. Чем больше лучей, тем числа их пересечения больше коррелированы между собой и тем больше количество измеряемых параметров. Слишком мало лучей брать тоже невыгодно, так как теряется часть информации. Эксперименты указывают (см. гл. 4), что рациональное количество лучей от 4 до 8.

Для определения квазитопологических инвариантов в некоторых случаях выгодно привлечь также теорию выбросов случайных процессов. Действительно, определение среднего числа пересечений линии связи, дисперсии числа пересечений и распределения точек пересечений сводится к известным в теории выбросов случайных функций задачам установления этих же статистических характеристик пересечений соответствующих уровней функциями $\varphi(t)$ или $z(t)$. При этом предполагается, что статистические свойства функций $\varphi(t)$ или $z(t)$ известны. Они могут быть определены на основании известных статистических свойств

исходного сигнала $s(t)$ и преобразований, приводящих к образованию структурных связей $\xi_1(t)$, $\xi_2(t)$ и самих функций $\varphi(t)$, $z(t)$. С методикой подобных расчетов можно познакомиться в соответствующей литературе, например [25]. К сожалению для речевых сигналов такие расчеты даже при помощи ЭВМ проделать довольно трудно, так как их статистические свойства недостаточно изучены.

Так, например, среднее число пересечений нулевого уровня или среднее число нулей N функции $z(t)$ за время T рассчитывается по формуле:

$$N = \int_{t_0}^{t_0+T} dt \int_{-\infty}^{\infty} |z(t)| W_2[0, \dot{z}(t)] dz,$$

где $W_2[z(t), \dot{z}(t)]$ — совместная плотность вероятности процесса $z(t)$ и его производной $\dot{z}(t)$ в совпадающие моменты времени. Дисперсия числа нулей

$$\sigma_N^2 = N - N^2 + \int_0^{\infty} \int_0^{\infty} F(t_1, t_2) dt_1 dt_2,$$

где функция $F(t_1, t_2)$ определяется выражением:

$$F(t_1, t_2) = \int_0^{\infty} \int_0^{\infty} |\dot{z}_1| |\dot{z}_2| W_4[0, \dot{z}_1, 0, \dot{z}_2] dz_1 dz_2.$$

Здесь $W_4[z(t_1), \dot{z}(t_1), z(t_2), \dot{z}(t_2)] = W_4[z_1, \dot{z}_1, z_2, \dot{z}_2]$ — совместная плотность вероятности процесса $z(t)$ и его производной $\dot{z}(t)$ в два разных момента времени.

Отметим, что в случае функции $\varphi(t)$ формулы для расчетов N и σ_N^2 были бы аналогичными. Таким образом, уже определение наипростейших характеристик N и σ_N^2 требует знания плотностей вероятности W_2 и W_4 для функции $z(t)$ или $\varphi(t)$, которые никак нельзя установить, если не известны соответствующие плотности вероятности для исходного речевого сигнала $s(t)$. Очевидно, определение подобных характеристик речевых сигналов требует большой экспериментальной работы, но несомненно может быть выполнено.

В качестве иллюстрации сказанного и для более глубокого понимания физической стороны вопроса рассмотрим несколько подробнее квантопологические характеристики для элементарного фазового изображения $X = \xi_1(t) = s(t)$, $Y = \xi_2(t) = \dot{s}(t)$ ($s(t)$ — речевой сигнал). Пусть линиями связи служат лучи связи (см. рис. 7). Тогда число пересечений оси ординат соответствует числу переходов через нуль речевого сигнала, а число пересечений оси абсцисс — числу его экстремумов (максимумов и минимумов). Это хорошо известные и широко используемые характеристики речевого сигнала. Вместе с тем очевидно, что числа пересечений лучей связи — более тонкие характеристики. Так, число пересечений первого луча соответствует числу экстремумов положительной полуволны сигнала, а число пересечений пятого луча — числу экстремумов отрицательной полуволны сигнала. Числа пересече-

ний третьего и седьмого лучей равны между собой, так как соответствуют числам переходов через нуль сигнала с положительным и отрицательным наклонами. Пересечения же остальных лучей — это переходы через нуль линейных комбинаций положительной или отрицательной полуволны сигнала и их производных.

В заключение этого параграфа рассмотрим еще одну возможность выбора линий связи, приводящую к дальнейшему расширению множества квазитопологических характеристик. Эта возможность заключается в подсчете чисел пересечений линий связи только в пределах некоторых их отрезков, определяемых

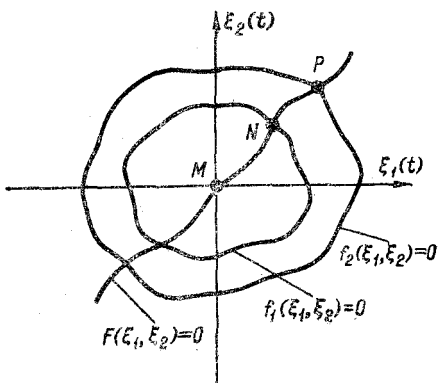


Рис. 9. Пересечения фазовым изображением отрезков линии связи

заданными значениями модуля радиус-вектора. В качестве примера такой случай изображен на рис. 9, где линия связи $F(\xi_1, \xi_2) = 0$ делится на отрезки двумя замкнутыми кривыми $f_1(\xi_1, \xi_2) = 0$ и $f_2(\xi_1, \xi_2) = 0$, определяющими заданные значения модуля радиус-вектора. Очевидно, функцию $z(t)$ в этом случае надо определять для каждого отрезка линии связи отдельно. Так, например, на отрезке MN функция $z(t)$ имеет вид: $z(t) = F(\xi_1, \xi_2)$, $f_1(\xi_1, \xi_2) < 0$, а на отрезке NP $z(t) = F(\xi_1, \xi_2)$, $f_1(\xi_1, \xi_2) > 0$, $f_2(\xi_1, \xi_2) < 0$. Аналогично могут быть записаны выражения $z(t)$ и на других отрезках.

При интерпретации пересечений фазовым изображением отрезков линии связи как пересечений функцией $\varphi(t)$ кривых $\varphi_1(t)$ и $\varphi_2(t)$ необходимо учитывать дополнительные условия, ограничивающие значения модуля радиус-вектора на рассматриваемых отрезках (см. рис. 9 и 10). Обозначая модуль радиус-вектора через $A(t)$ (очевидно, $A(t) = \sqrt{\xi_1^2(t) + \xi_2^2(t)}$), а кривые, определяющие его значения, через $A_1(t)$ и $A_2(t)$ (на фазовой плоскости им соответствуют кривые $f_1(\xi_1, \xi_2) = 0$, $f_2(\xi_1, \xi_2) = 0$), эти условия

можно записать в виде: $A(t) < A_1(t)$, $A_1(t) < A(t) < A_2(t)$, $A(t) > A_2(t)$ в зависимости от рассматриваемого отрезка). Так, в примере на рис. 10 учитывается условие $A_1(t) < A(t) < A_2(t)$, что сводится к подсчету только тех пересечений $\varphi(t)$ с $\varphi_1(t)$ и $\varphi_2(t)$, которые во времени совпадают с выполнением этого условия (они обозначены более толстой штриховой).

Удобно в качестве кривых $f_1(\xi_1, \xi_2) = 0$, $f_2(\xi_1, \xi_2) = 0$ использовать круги с центрами в начале координат. Тогда $A_1(t)$ и $A_2(t)$ становятся прямыми, параллельными оси времени и проходящими на расстоянии от нее, равном радиусу соответствующего круга на фазовой плоскости.

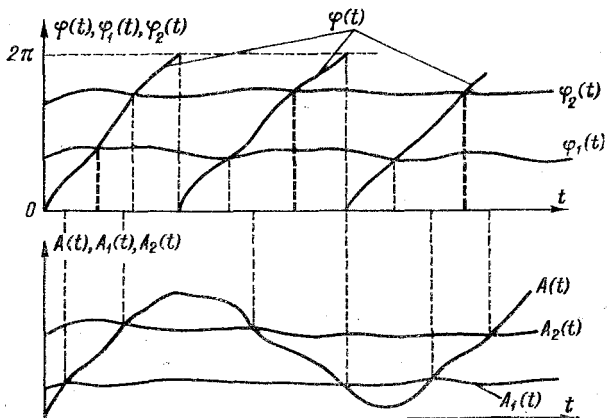


Рис. 10. Интерпретация пересечений фазовым изображением отрезков линии связи

В терминах теории выбросов подсчет чисел пересечений отрезков линий связи сводится к задаче определения статистических свойств выбросов одной функции при условии одновременного существования выбросов другой функции. Однако в такой постановке задача является мало исследованной.

8. Время пребывания изображающей точки фазового изображения в секторах плоскости

К возможным интегральным мерам фазовых изображений относится также время вращения радиального поля. Это время, в течение которого изображающая точка фазового изображения находится в определенных секторах плоскости. Очевидно, эта характеристика связана со скоростью и направлением движения изображающей точки и может быть определенным образом выражена через угол поворота радиус-вектора относительно по-

люса. Так, при параметрическом задании фазовой траектории через $\xi_1(t)$ и $\xi_2(t)$ это время определится соотношением

$$\tau = \int_{t_0}^{t_0+T} \eta(t) dt, \quad (12)$$

где

$$\eta(t) = \frac{1}{2} \{ \text{sign} [\varphi(t) - \varphi_1(t)] - \text{sign} [\varphi(t) - \varphi_2(t)] \}. \quad (13)$$

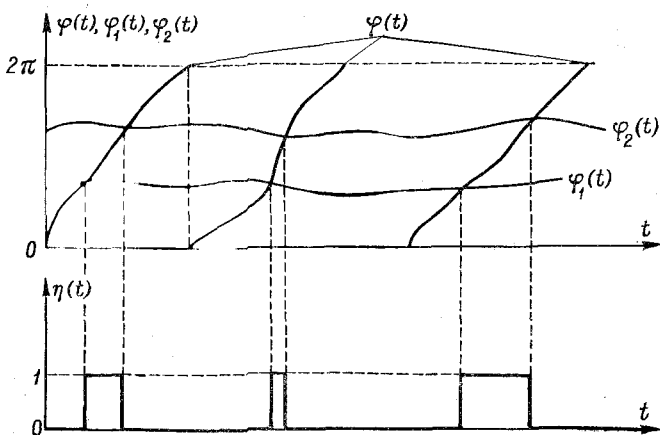


Рис. 11. Образование меандровой функции $\eta(t)$

Здесь $\varphi(t)$ — мгновенное значение угла поворота радиус-вектора изображающей точки (см. (9)); $\varphi_1(t)$, $\varphi_2(t)$ — углы, определяющие границы сектора (при этом должно быть соблюдено условие $\varphi_2(t) > \varphi_1(t)$); T — время наблюдения сигнала; знаковая функция

$$\text{sign } x = \begin{cases} 1, & x > 0; \\ -1, & x < 0. \end{cases}$$

Для более глубокого понимания физического смысла выражений (12) и (13) приведен рис. 11, где иллюстративно изображены функции $\varphi(t)$, $\varphi_1(t)$, $\varphi_2(t)$ и $\eta(t)$, соответствующие одному сектору на фазовой плоскости.

Задание углов $\varphi_1(t)$, $\varphi_2(t)$ равносильно заданию граничных линий сектора. Показанная на рис. 12 плоскость при помощи граничных линий разделена на четыре сектора. Если уравнения

границных линий секторов заданы в виде $z_1(t) = F_1(\xi_1, \xi_2) = 0$ и $z_2(t) = F_2(\xi_1, \xi_2) = 0$, то функция $\eta(t)$ из выражения (13) (можно ее назвать меандровой функцией) приобретает вид:

для первого сектора

$$\eta(t) = \eta_1(t) = \begin{cases} 1 & \text{при } F_1(\xi_1, \xi_2) > 0, F_2(\xi_1, \xi_2) < 0; \\ 0 & \text{во всех других случаях;} \end{cases}$$

для второго сектора

$$\eta(t) = \eta_2(t) = \begin{cases} 1 & \text{при } F_1(\xi_1, \xi_2) > 0, F_2(\xi_1, \xi_2) > 0; \\ 0 & \text{во всех других случаях;} \end{cases}$$

для третьего сектора

$$\eta(t) = \eta_3(t) = \begin{cases} 1 & \text{при } F_1(\xi_1, \xi_2) < 0, F_2(\xi_1, \xi_2) > 0; \\ 0 & \text{во всех других случаях;} \end{cases}$$

для четвертого сектора

$$\eta(t) = \eta_4(t) = \begin{cases} 1 & \text{при } F_1(\xi_1, \xi_2) < 0, F_2(\xi_1, \xi_2) < 0; \\ 0 & \text{во всех других случаях.} \end{cases}$$

(14)

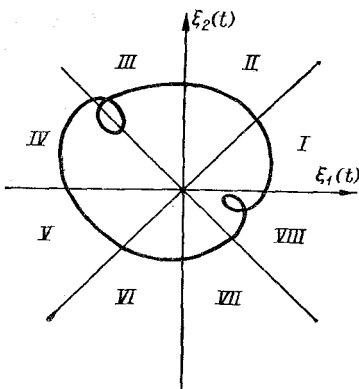
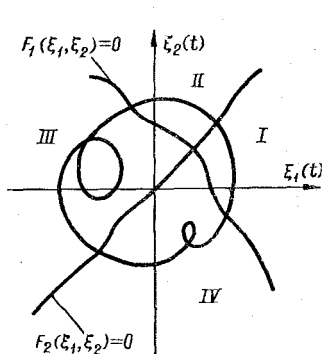


Рис. 12. Разбиение фазовой плоскости на сектора при помощи граничных линий

Рис. 13. Разбиение фазовой плоскости на сектора при помощи прямых граничных линий

Как и в случае квазитопологических характеристик, целесообразным является переход к прямым граничным линиям секторов. Уравнения линий принимают вид: $z(t) = \xi_2(t) \pm k\xi_1(t)$, а функции $\varphi_1(t), \varphi_2(t) \dots$ становятся константами. Следуя (14),

легко записать выражение функции $\eta(t)$ в этом случае. На рис. 13, например:

$$\left. \begin{aligned}
 \eta_1(t) &= \begin{cases} 1 & \text{при } \xi_2(t) - k\xi_1(t) < 0, \quad \xi_2(t) > 0; \\ 0 & \text{во всех других случаях} \end{cases} \\
 \eta_2(t) &= \begin{cases} 1 & \text{при } \xi_2(t) - k\xi_1(t) > 0, \quad \xi_1(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_3(t) &= \begin{cases} 1 & \text{при } \xi_2(t) + k\xi_1(t) > 0, \quad \xi_1(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_4(t) &= \begin{cases} 1 & \text{при } \xi_2(t) + k\xi_1(t) < 0, \quad \xi_2(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_5(t) &= \begin{cases} 1 & \text{при } \xi_2(t) - k\xi_1(t) > 0, \quad \xi_2(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_6(t) &= \begin{cases} 1 & \text{при } \xi_2(t) - k\xi_1(t) < 0, \quad \xi_1(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_7(t) &= \begin{cases} 1 & \text{при } \xi_2(t) + k\xi_1(t) < 0, \quad \xi_2(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_8(t) &= \begin{cases} 1 & \text{при } \xi_2(t) + k\xi_1(t) > 0, \quad \xi_2(t) < 0; \\ 0 & \text{во всех других случаях.} \end{cases}
 \end{aligned} \right\} \quad (15)$$

Выбор числа секторов составляет отдельную задачу, удовлетворительного решения которой в общем виде пока нет. Можно высказать лишь соображения, подобные приведенным в предыдущем параграфе при выборе числа лучей связи.

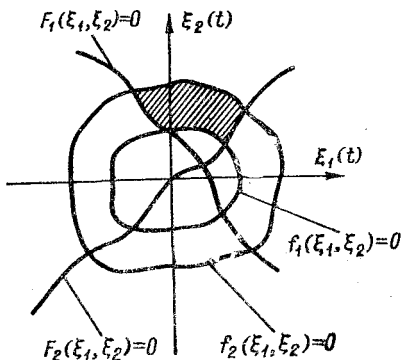


Рис. 14. Разбиение фазовой плоскости на сектора по модулю и по углу поворота радиус-вектора

при решении подобных задач приходится столкнуться с большими вычислительными трудностями, а также с недостатком априорной информации о статистических свойствах речевых сигналов.

В качестве иллюстрации рассмотрим опять же фазовое изображение вида $\xi_1(t) = s(t)$, $\xi_2(t) = \dot{s}(t)$ где $s(t)$ — речевой сигнал (см. рис. 13). Время пребывания изображающей точки в правой полуплоскости — это суммарная длительность положительных выбросов сигнала $s(t)$ над нулевым уровнем, время пребывания в левой полуплоскости — длительность отрицательных выбросов. Время пребывания в верхней полуплоскости связано с длительностью положительных выбросов производной речевого сигнала, а время пребывания в нижней полуплоскости —

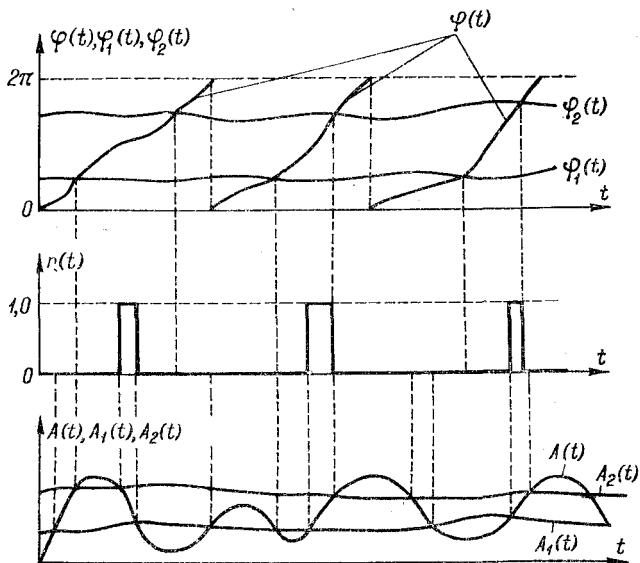


Рис. 15. Образование меандровой функции $\eta(t)$, определяемой выбросами модуля и угла поворота радиус-вектора

с длительностью отрицательных выбросов производной. При делении плоскости на большее число секторов получаются более тонкие характеристики. Так, например, в случае восьми секторов, время пребывания в первом из них (рис. 13) — это длительность отрицательных выбросов функции $\dot{s}(t) - ks(t)$ при условии $\dot{s}(t) > 0$.

Как и в случае определения квазитопологических характеристик, формирование секторов плоскости, в которых подсчитывается время пребывания, можно осуществлять не только по углу поворота радиус-вектора, но и по его модулю (рис. 14). Тогда выражение меандровой функции $\eta(t)$, например для заштрихованного сектора на рис. 14, имеет вид:

$$\eta(t) = \begin{cases} 1 & \text{при } F_1(\xi_1, \xi_2) \triangleright 0, F_2(\xi_1, \xi_2) \triangleright 0, \\ & f_1(\xi_1, \xi_2) \triangleright 0, f_1(\xi_1, \xi_2) \triangleleft 0; \\ 0 & \text{во всех других случаях.} \end{cases}$$

Для других секторов выражения функций $\eta(t)$ могут быть записаны аналогично.

Очевидно, при таком формировании секторов меандровая функция $\eta(t)$ определяется выбросами не только функции $\varphi(t)$ (мгновенного значения угла поворота радиус-вектора), но и функции $A(t)$ (модуля радиус-вектора). В качестве иллюстрации на рис. 15 показано формирование функции $\eta(t)$ для одного сектора плоскости (для сравнения см. также рис. 11).

9. Кодирование фазовых изображений

Кодирование фазовых изображений тесно связано с определением квазитопологических характеристик, а также времени пребывания изображающей точки в секторах плоскости и является дальнейшим развитием этих двух методов извлечения полезной информации на основании использования фазовых изображений. Метод кодирования предполагает разбиение фазовой плоскости на ряд секторов при помощи граничных линий или только по углу поворота радиус-вектора, или одновременно и по его модулю (см. рис. 12, 14) и присвоение этим секторам некоторых кодовых комбинаций. Тогда фазовое изображение характеризуется последовательностью этих кодовых комбинаций, получаемых при движении изображающей точки через соответствующие секторы. При таком описании фазовых изображений уже учитывается их динамика, в частности направление вращения радиус-вектора и изменение его модуля. Естественно, при этом еще больше сужается круг допустимых преобразований изображений, относительно которых достигается инвариантность, по сравнению с преобразованиями при использовании квазитопологических характеристик или времени пребывания.

Самостоятельный круг вопросов связан с выбором самих кодовых комбинаций. Очевидно, с целью избежания значительного воздействия помех, переводящих одну кодовую комбинацию в другую, соприкасающиеся секторы следует закодировать близкими (по расстоянию Хемминга) кодовыми комбинациями, а противоположные — дальними. Однако более детальное рассмотрение подобных вопросов выходит за рамки настоящей книги.

Наиболее целесообразным с точки зрения простоты реализации кодирующего устройства является использование двоичного кода. При этом количество разрядов кода m и число секторов n связываются формулой $m = \log_2 n$. Так что чрезмерное увеличение числа секторов приводит к значительному увеличению разрядности кода, а это, в свою очередь, вызывает затруднения при дальнейшей обработке кодовых комбинаций.

10. Интегральные оценки структурных связей с явным временем

Среди числовых характеристик структурных связей с явным временем наиболее целесообразным, на наш взгляд, является использование их интегральных оценок. Интегральные оценки

структурных связей с явным временем можно рассматривать как результат их определенной интегральной обработки. Операцию указанной обработки естественно записать в виде:

$$J_{\xi} = \int_{t_0}^{t_0+T} \xi(t) dt = \int_{t_0}^{t_0+T} \Phi \{L_1[s(t)], L_2[s(t)], \dots\} dt. \quad (16)$$

Особую роль интегральные оценки приобретают тогда, когда они берутся с некоторым «весом», т. е.

$$J'_{\xi} = \int_{t_0}^{t_0+T} \mu(t) \xi(t) dt. \quad (17)$$

При соответствующем выборе «веса» $\mu(t)$ удастся подчеркнуть (либо нивелировать) влияние на результаты анализа тех или иных свойств рассматриваемых структурных связей, т. е. перейти к синтезу алгоритмов по определенному критерию качества работы. Функция $\mu(t)$ может включать в себя в общем случае ряд неопределенных коэффициентов, которыми целесообразно распорядиться, исходя из принятого критерия качества. Отметим, что эти вопросы пока мало изучены; более детальное их рассмотрение намечается.

Рассмотрим несколько подробнее некоторые свойства интегральных оценок структурных связей, введенных в § 3. Это рассмотрение базируется на возможности аналитического выполнения интегрирования согласно (16) для некоторых видов структурных связей.

Так, интегральная оценка диссипанта равна:

$$J_{\delta} = \int_{t_0}^{t_0+T} \delta(t) dt = \int_{t_0}^{t_0+T} \frac{\dot{x}(t)}{x(t)} dt = \ln[x(t)] \Big|_{t_0}^{t_0+T} = \ln \frac{x(t_0+T)}{x(t_0)}.$$

В соответствии с обычным определением информации оценка J_{δ} в этом случае представляет собой приращение количества информации на интервале T в момент времени t_0 при условии, что пороговый уровень равен $x(t_0)$.

Когда в качестве контурной модели $x(t)$ применяются спектральные огибающие речевого сигнала, интегральная оценка диссипанта равна значению фонетической функции Пирогова при $t=t_0$ (см. § 4).

В случае кондисанта будем иметь:

$$J_{\rho} = \int_{t_0}^{t_0+T} \rho(t) dt = \int_{t_0}^{t_0+T} \frac{\ddot{x}(t)}{\dot{x}(t)} dt = \ln[\dot{x}(t)] \Big|_{t_0}^{t_0+T} = \ln \frac{\dot{x}(t_0+T)}{\dot{x}(t_0)}.$$

Нетрудно заметить, что J_{ρ} совпадает с J_{δ} в случае, когда J_{δ} рассматривается для первой производной контурной модели $x(t)$.

Для кондисквadrата получаем:

$$J_{\lambda} = \int_{t_0}^{t_0+T} \lambda(t) dt = \int_{t_0}^{t_0+T} \frac{x(t) \ddot{x}(t)}{\dot{x}^2(t)} dt = \\ = t \left|_{t_0}^{t_0+T} - \frac{x(t)}{\dot{x}(t)} \right|_{t_0}^{t_0+T} = T - \left[\frac{1}{\delta(t_0)} - \frac{1}{\delta(t_0+T)} \right],$$

т. е. интегральная оценка кондисквadrата выражается через постоянную составляющую, зависящую от времени наблюдения сигнала, и разницу значений функции обратной диссипанту в начальный и в конечный моменты наблюдения.

11. Некоторые вопросы предварительной сегментации речевых сигналов

До сих пор мы сознательно обходили вопрос об установлении необходимого времени наблюдения сигнала (оно обозначено через T в формулах (6), (7), (12), (16), (17)), определяющего те его отрезки, для которых целесообразно проводить измерение рассмотренных выше информативных параметров: квазитопологических характеристик, времени пребывания, интегральных оценок. Естественно, длительность этих отрезков должна быть меньше или по крайней мере равна длительности основных единиц распознавания, в качестве которых, как правило, выбирают слова, слоги или отдельные фонемы. С уменьшением длительности отрезков увеличивается объем перерабатываемой информации, в чем нет никакой необходимости в большинстве случаев при решении задач распознавания небольшого набора слов или фонем. С увеличением же числа распознаваемых единиц или с повышением требований надежности распознавания целесообразно уменьшать длительность отрезков для получения более полного, более информативного описания.

Границы между соседними отрезками в принципе можно ставить чисто формально, например через равные интервалы. Такой подход имеет свои преимущества, особенно в смысле простоты аппаратной реализации. Однако это совершенно не соответствует известным данным о физической природе речевого сигнала и о механизмах слухового восприятия. Кроме того, рассмотренные нами информативные параметры являются интегральными параметрами, т. е. отражают не мгновенные, а усредненные за время T свойства сигнала. В связи с этим, целесообразно измерение параметров проводить на отрезках такой длительности, при которой свойства сигнала не успевают много измениться. Следовательно, в такой постановке задача определения границ отрезков превращается в задачу предварительной сегментации речевых сигналов. (Не надо отождествлять это с сегментацией вообще, под которой подразумевается объективное разделение речевого сообщения на отрезки, соотносимые с рассматриваемыми основными единицами, в большинстве слу-

чаев с фонемами.) Ниже рассматриваются некоторые свойства отрезков речевых сигналов и способы их предварительной сегментации на основании структурных методов.

Некоторые свойства отрезков речевых сигналов. Непрерывная речь образуется как последовательность отрезков сигнала. Однако границы между этими отрезками установить очень трудно, так как речевой тракт в силу своей инертности не может образовать ступенчатые переходы от одного отрезка к другому. В действительности эти переходы осуществляются плавно и, кроме того, соседние отрезки взаимодействуют между собой (эффект коартикуляции), как показано на рис. 16.

Отрезки сигнала могут иметь начальную переходную часть, стационарную часть и конечную переходную часть. При пере-

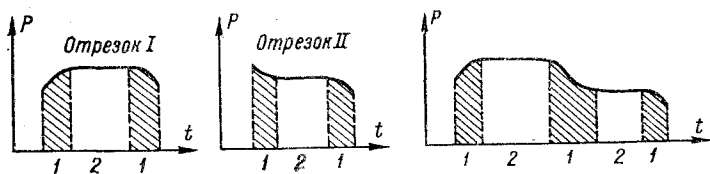


Рис. 16. Примеры отрезков речевых сигналов и их взаимодействия
 P — параметр, 1 — переходная часть, 2 — стационарная часть

ходе от одного отрезка к другому конечная переходная часть одного отрезка и начальная переходная часть другого отрезка сливаются в динамический переходный участок. Следует подчеркнуть, что большинство речевых отрезков имеют стационарную часть, длительность которой существенно зависит от темпа речи и от индивидуальных особенностей диктора. Такие отрезки встречаются в гласных, назальных, фрикативных и полугласных звуках. Отметим, что путем объединения стационарных и переходных отрезков с одновременным использованием взаимоотношений между ними, лингвистической информации и результатов распознавания можно однозначно образовать сегменты сигнала, большей частью сопоставимые с фонемами. Однако эти вопросы уже выходят за рамки настоящей работы, и мы не будем их касаться.

Способы выделения стационарных и переходных отрезков. Идея преобразований речевого сигнала, приводящих к образованию его отрезков, заключается в обнаружении моментов изменения энергии в частотных каналах. Наиболее успешно эту задачу можно решать, применяя модель адаптивного механизма (рис. 2), вкратце рассмотренную в § 4, или, используя структурные связи: диссипант, кондисант и их интегральные оценки.

Действительно, анализ работы модели адаптивного механизма как самонастраивающейся системы [17] указывает на то, что в любом частотном канале на его выходе появляются

скачки выходного сигнала в моменты быстрого изменения энергии в соответствующем участке частотного спектра. Следовательно, отсутствие таких скачков во всех каналах будет свидетельствовать о стационарности сигнала, а появление их — о переходном участке. В связи с тем что в действительности строгой стационарности в речевых сигналах нет, для более надежной сегментации решение следует принимать не по срабатыванию одного адаптивного элемента, а при обнаружении переходов в нескольких частотных каналах одновременно.

Аналогично дело обстоит и при использовании интегральных оценок диссипанта и конденсанта для огибающих в частотных каналах. Соответствующие выражения (см. § 10, 4) указывают на то, что эти оценки не претерпевают существенных изменений при медленных вариациях огибающих и, наоборот, изменяются сильно при быстрых их вариациях. На стационарных же отрезках звуков оценки приближаются к нулю. Очевидно, эти свойства могут быть положены в основу работы устройств предварительной сегментации речевых сигналов.

Глава третья

ВЫДЕЛЕНИЕ И ИЗМЕРЕНИЕ ПРИЗНАКОВ

12. Измерение квазитопологических характеристик

Согласно определению (см. § 7) квазитопологические характеристики — это числа нулей некоторой функции $z(t)$, представляющей собой определенную связь между фазовыми координатами фазового изображения. Это и положено в основу принципа действия соответствующих устройств выделения. В самом общем случае структурная схема такого устройства имеет вид, изображенный на рис. 17¹. Речевой сигнал $s(t)$ из микрофона или магнитофона после усиления в предварительном усилителе 1 поступает на вход функциональных преобразователей 2 и 3, формирующих фазовые координаты $\xi_1(t)$ и $\xi_2(t)$ соответствующего фазового изображения. Далее следует функциональный преобразователь 4, формирующий функцию $z(t)$ согласно выражению (8). Выделение нулей этой функции осуществляется с использованием принципа клиппирования.

Для этого служит ограничитель 10, на двух выходах которого положительные и отрицательные полуволны функции $z(t)$ представляются однополярными прямоугольными импульсами. (Знаком плюс обозначен выход, соответствующий положитель-

¹ На этом и на других рисунках этой главы нумерация отдельных блоков структурных схем не сплошная. Это сделано с той целью, чтобы одинаковые функциональные блоки на разных рисунках обозначались одними номерами.

ной полувольтной функции $z(t)$, а знаком минус — отрицательной полувольтной). Затем следуют дифференцирующие RC -цепочки 11, выходные короткие импульсы которых (одной полярности) складываются в одну последовательность во времени при помощи схемы 12, осуществляющей логическую операцию ИЛИ. Очевидно, появляющаяся на выходе последовательность коротких импульсов во времени совпадает с нулями функции $z(t)$. Их подсчет может быть осуществлен при помощи двоичного счетчика.

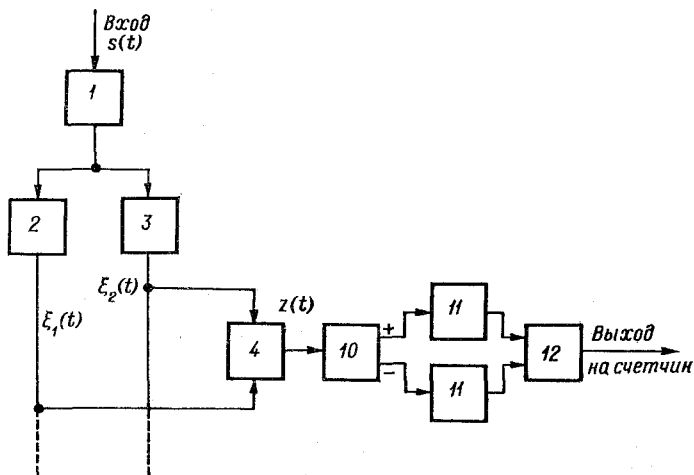


Рис. 17. Структурная схема устройства для подсчета числа пересечений

С технической точки зрения наиболее сложными блоками в рассмотренной схеме, очевидно, являются нелинейные функциональные преобразователи 2, 3, 4. Не исключено, что в некоторых случаях осуществление этих преобразователей с заданными метрологическими характеристиками может встретить большие затруднения. Однако во многих практических случаях, когда в качестве линий связи используются лучи связи, а в качестве фазовых координат — произведения и частные производных разных порядков контурных моделей исходного речевого сигнала, можно построить более простые схемы, содержащие лишь перемножители вместо нелинейных функциональных преобразователей, осуществление же перемножителей в настоящее время не является проблемой [23]. Естественно, нельзя обойтись без преобразователей для формирования самих контурных моделей, за исключением случаев, когда в качестве контурной модели используется сам исходный сигнал. Действительно, нами предложенные в § 3 фазовые изображения содержат фазовые координаты, выражаемые только через произведения и частные

производных контурных моделей. Нетрудно показать, что функция $z(t)$ для подобных фазовых изображений, пересекаемых лучами связи, может быть выражена только через произведения и суммы производных контурных моделей сигнала. Другими словами для формирования функции $z(t)$ в этом случае необхо-

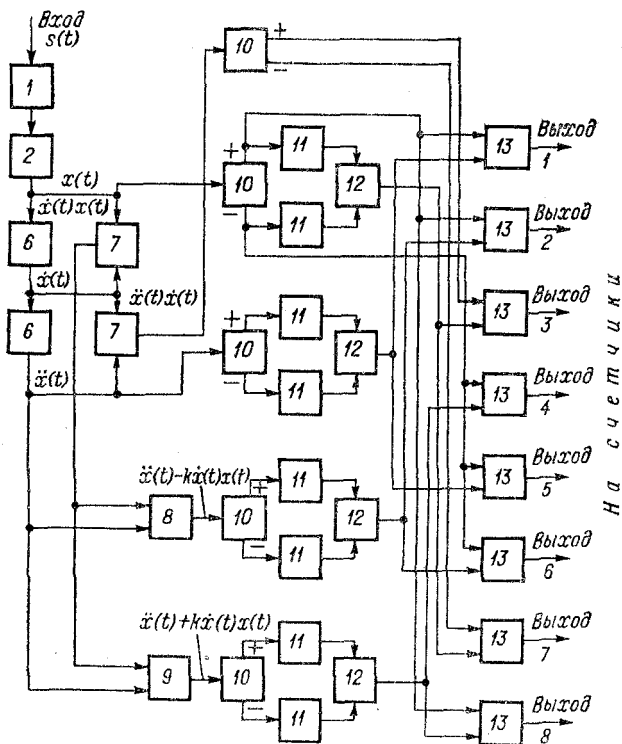


Рис. 18. Структурная схема устройства для подсчета числа пересечений на плоскости $x(t)$, $\rho(t)$

димы дифференциаторы, перемножители и сумматоры (естественно, при этом предполагается, что контурные модели сформированы отдельно).

В качестве иллюстрации сказанного запишем выражение функции $z(t)$ в случае фазового изображения $z(t)$, $\rho(t) = \frac{\ddot{x}(t)}{x(t)}$, пересекаемого восьмью лучами связи (см. рис. 7 и соотношения (11)):

для первого луча	$z(t) = \ddot{x}(t), \quad x(t) > 0;$	} (18)
для второго луча	$z(t) = \ddot{x}(t) - k\dot{x}(t)x(t), \quad x(t) > 0;$	
для третьего луча	$z(t) = x(t), \quad \ddot{x}(t)\dot{x}(t) > 0;$	
для четвертого луча	$z(t) = \ddot{x}(t) + k\dot{x}(t)x(t), \quad x(t) < 0;$	
для пятого луча	$z(t) = \dot{x}(t), \quad x(t) < 0;$	
для шестого луча	$z(t) = \ddot{x}(t) - k\dot{x}(t)x(t), \quad x(t) < 0;$	
для седьмого луча	$z(t) = x(t), \quad \ddot{x}(t)\dot{x}(t) < 0;$	
для восьмого луча	$z(t) = \dot{x}(t) + k\dot{x}(t)x(t), \quad x(t) > 0.$	

На рис. 18 представлена структурная схема устройства выделения квазитопологических характеристик для рассматриваемого случая, составленная согласно выражениям (18). Речевой сигнал $s(t)$ через усилитель 1 поступает на преобразователь 2

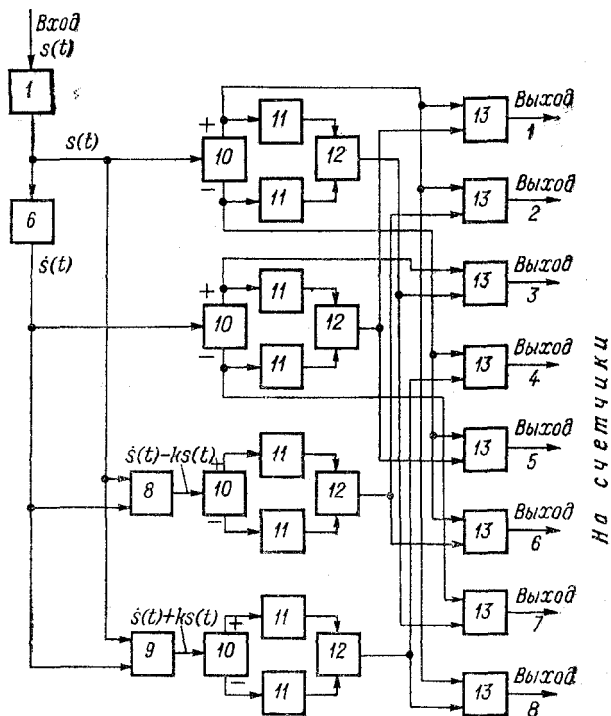


Рис. 19. Структурная схема устройства для подсчета числа пересечений на элементарной плоскости $s(t), \dot{s}(t)$

для формирования контурной модели $x(t)$. Далее следуют два одинаковых дифференциатора 6, два одинаковых переключателя 7 и два сумматора 8 и 9, первый из которых в действительности осуществляет операцию вычитания, а второй — непосредственного суммирования. Ограничители 10, дифференцирующие цепочки 11 и схемы ИЛИ 12 служат для формирования коротких импульсов, совпадающих по времени с нулями функций $x(t)$, $\dot{x}(t)$, $\dot{x}(t) + kx(t)x(t)$ и $\dot{x}(t) - kx(t)x(t)$. Блоки 13 — схемы, выполняющие логическую операцию И. Они пропускают на выход устройства только те импульсы, которые совпадают во времени с выполнением определенных условий (18) для соответствующих лучей. Таким образом, на выходах 1, 2, ..., 8 устройства получаем последовательности импульсов, количество которых совпадает с числом пересечений фазовым изображением \dot{x}/x , x соответствующих лучей связи.

Аналогичным образом можно построить структурные схемы устройств для выделения квазитопологических характеристик и в случае других фазовых изображений.

В заключение приведем структурную схему устройства выделения квазитопологических характеристик на элементарной фазовой плоскости $s(t)$, $\dot{s}(t)$, пересекаемой восьмью лучами связи (см. рис. 7). (Согласно нами принятым обозначениям $\xi_1(t) = s(t)$, $\xi_2(t) = \dot{s}(t)$, где $s(t)$ — исходный речевой сигнал.) Схема изображена на рис. 19, а выражения функций $z(t)$, по которым построена схема, приведены ниже:

для первого луча	$z(t) = \dot{s}(t), \quad s(t) > 0;$	}	(19)
для второго луча	$z(t) = \dot{s}(t) - ks(t), \quad s(t) > 0;$		
для третьего луча	$z(t) = s(t), \quad \dot{s}(t) > 0;$		
для четвертого луча	$z(t) = \dot{s}(t) + ks(t), \quad s(t) < 0;$		
для пятого луча	$z(t) = s(t), \quad s(t) < 0;$		
для шестого луча	$z(t) = \dot{s}(t) - ks(t), \quad s(t) < 0;$		
для седьмого луча	$z(t) = s(t), \quad \dot{s}(t) < 0;$		
для восьмого луча	$z(t) = \dot{s}(t) + ks(t), \quad s(t) > 0.$		

13. Измерение времени пребывания

Перейдем теперь к рассмотрению способов выделения признаков, основанных на измерении времени пребывания изображающей точки в секторах фазового изображения.

В общем случае (см. § 8) речь идет о любых функционально-преобразованных фазовых изображениях, а также о любого вида граничных линиях секторов, в которых измеряется время пребывания изображающей точки.

Идея алгоритма измерения времени пребывания заложена в выражениях (12), (13), а также (14) и заключается в формировании меандровой функции $\eta(t)$ с последующим ее интегри-

рованием. Однако более удобно и выгодно с точки зрения точности измерения заменить непосредственное интегрирование подсчетом количества импульсов некоторого стабильного тактового генератора.

Структурная схема такого устройства приведена на рис. 20 (изображена только часть схемы для измерения времени пребывания в первом секторе согласно рис. 12 и соотношениям (14)). Речевой сигнал $s(t)$ после усиления в предварительном усилителе 1 поступает на два функциональных преобразователя 2 и

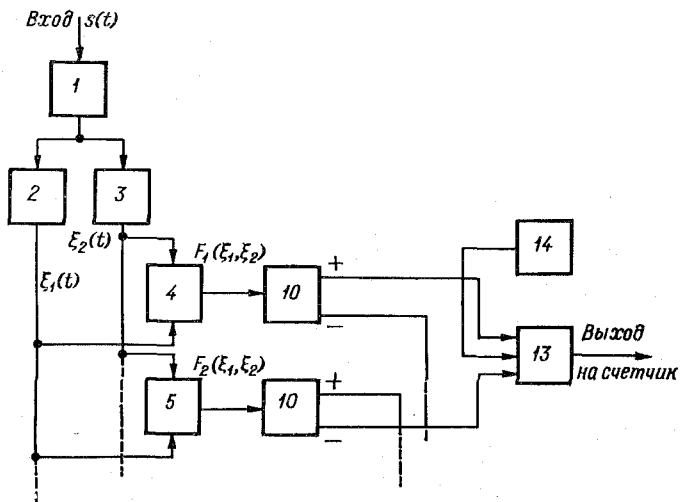


Рис. 20. Структурная схема устройства для измерения времени пребывания

3, формирующих фазовые координаты $\xi_1(t)$ и $\xi_2(t)$ соответствующего фазового изображения. Функциональные преобразователи 4 и 5 формируют определенные связи между фазовыми координатами $F_1(\xi_1, \xi_2)$ и $F_2(\xi_1, \xi_2)$, соответствующие уравнениям граничных линий секторов плоскости. Ограничители 10 клиппируют функции $F_1(\xi_1, \xi_2)$, $F_2(\xi_1, \xi_2)$, выдавая на выходах однополярные прямоугольные импульсы, свидетельствующие о выполнении условий существования функции $\eta(t)$ — см. (14). При одновременном выполнении соответствующих условий (в данном случае условий для первого сектора) на выход через схему 13 проходят короткие импульсы тактового генератора 14. Таким образом, в нашем случае количество выходных импульсов пропорционально длительности существования функции $\eta_1(t)$ (см. выражение (14)), а тем самым и времени пребывания изображающей точки в первом секторе плоскости $\xi_1(t)$, $\xi_2(t)$. Подсчет числа выходных импульсов можно осуществить при помощи двоичного счетчика.

Как и в случае измерения квантопологических инвариантов, наиболее сложными с реализационных позиций являются функциональные преобразователи. Однако, если фазовые координаты являются произведениями или частными производных контурной модели сигнала, а граничные линии секторов — прямыми линиями, можно обойтись вместо нелинейных функциональных преобразователей одними перемножителями. Действительно, в подобных случаях условия существования функции $\eta(t)$ выражаются только через произведения и суммы производных контурных моделей. Например, в случае плоскости $\xi_1(t) = x(t)$, $\xi_2(t) = \frac{\ddot{x}(t)}{\dot{x}(t)}$, разделенной прямыми линиями на 8 секторов (см. рис. 13 и выражения (15)), имеем:

$$\left. \begin{aligned} \eta_1(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) - k\dot{x}(t)x(t) < 0, \quad \ddot{x}(t)\dot{x}(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_2(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) - k\dot{x}(t)x(t) > 0, \quad x(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_3(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) + k\dot{x}(t)x(t) > 0, \quad x(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_4(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) + k\dot{x}(t)x(t) < 0, \quad \ddot{x}(t)\dot{x}(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_5(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) - k\dot{x}(t)x(t) > 0; \quad \ddot{x}(t)\dot{x}(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_6(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) - k\dot{x}(t)x(t) < 0, \quad x(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_7(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) + k\dot{x}(t)x(t) < 0, \quad x(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\ \eta_8(t) &= \begin{cases} 1 & \text{при } \ddot{x}(t) + k\dot{x}(t)x(t) > 0, \quad \ddot{x}(t)\dot{x}(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \end{aligned} \right\} (20)$$

Соответствующая структурная схема устройства приведена на рис. 21. Функции $x(t)$, $\dot{x}(t)x(t)$, $\ddot{x}(t) - k\dot{x}(t)x(t)$, $\ddot{x}(t)\dot{x}(t)$ формируются таким же образом, как и в схеме на рис. 18. После клиппирования соответствующих функций ограничителями 10, схемами И 13 проверяются условия существования функций $\eta(t)$ соответствующих секторов и регулируется подача импульсов тактового генератора 14 на выходы. Таким образом, на выходах 1, 2, ..., 8 появляются импульсы, количество которых пропорционально времени пребывания изображающей точки в секторах плоскости \dot{x}/x , x .

Подобным же образом строятся схемы устройств для измерения времени пребывания и в случае других фазовых плоскостей.

Для элементарного фазового изображения ($\xi_1(t) = s(t)$, $\xi_2(t) = \dot{s}(t)$), разделенного на восемь секторов прямыми линиями (рис. 13) будем иметь (см. также (15)):

$$\left. \begin{aligned}
 \eta_1(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) - ks(t) < 0, \quad \dot{s}(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_2(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) - ks(t) > 0, \quad s(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_3(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) + ks(t) > 0, \quad s(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_4(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) + ks(t) < 0, \quad \dot{s}(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_5(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) - ks(t) > 0, \quad \dot{s}(t) < 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_6(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) - ks(t) < 0; \quad s(t) < 0; \\ 0 & \text{во всех других случаях} \end{cases} \\
 \eta_7(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) + ks(t) < 0, \quad s(t) > 0; \\ 0 & \text{во всех других случаях;} \end{cases} \\
 \eta_8(t) &= \begin{cases} 1 & \text{при } \dot{s}(t) - ks(t) > 0, \quad \dot{s}(t) < 0; \\ 0 & \text{во всех других случаях.} \end{cases}
 \end{aligned} \right\} \quad (21)$$

Из эквивалентности функций $\dot{s}(t) - ks(t)$ и $\ddot{x}(t) - k\dot{x}(t)x(t)$, $\dot{s}(t) + ks(t)$ и $\dot{x}(t) + kx(t)x(t)$, $\dot{x}(t)x(t)$ и $\dot{s}(t)$, $x(t)$ и $s(t)$ в выражениях (21) и (20) следует эквивалентность схем устройств в случае плоскостей s , \dot{s} и x , ρ . Поэтому нет необходимости приводить схему устройства в данном случае. Укажем лишь, что функции $\dot{s}(t) - ks(t)$, $\dot{s}(t) + ks(t)$, $\dot{s}(t)$, $s(t)$ формируются так же, как в схеме на рис. 19, а дальнейшая обработка происходит, как в схеме на рис. 21.

14. Образование кодов фазовых изображений

Как уже было указано в § 9, принцип кодирования фазовых изображений основан на использовании информации о пребывании изображающей точки в секторах плоскости, т. е. той же самой информации, которая используется при измерении времени пребывания. В связи с этим структурные схемы устройств измерения времени пребывания и образования кодов во многом совпадают. Наглядно это видно на примере кодирования фазового изображения, приведенного на рис. 12. Структурная схема соответствующего устройства изображена на рис. 22. Показана только часть схемы, соответствующая кодированию фазового изображения в первом секторе, так как для других секторов схема аналогична. Формирование условий существования

меандровой функции $\eta_1(t)$, соответствующей пребыванию изображающей точки в первом секторе (см. выражения (14)), осуществляется так же, как и в схеме на рис. 20. При выполнении этих условий появляется импульс на выходе схемы И 13 и срабатывает кодирующее устройство 15, выдающее на выход

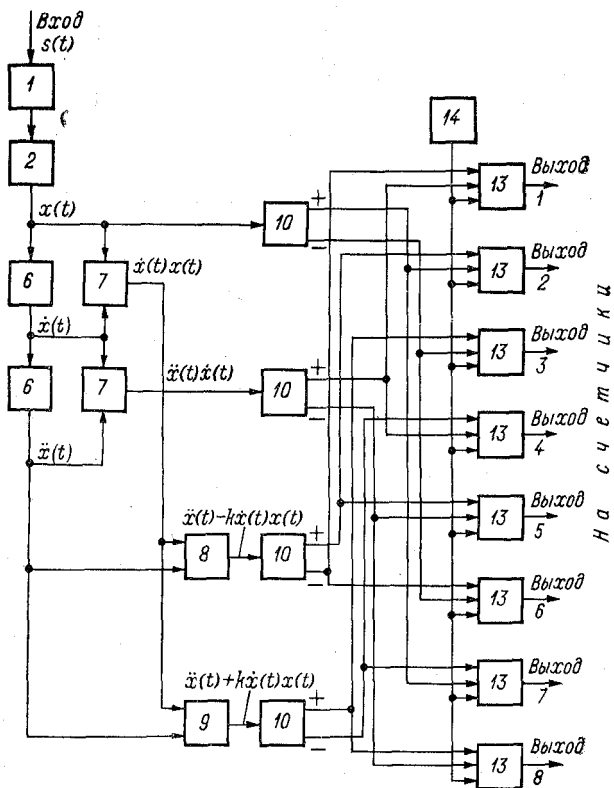


Рис. 21. Структурная схема устройства для измерения времени пребывания на плоскости $x(t)$, $\rho(t)$

заранее определенную m -разрядную кодовую комбинацию. Входы 2, 3, ..., n кодирующего устройства служат для подачи импульсов, соответствующих пребыванию изображающей точки в других секторах плоскости. При этом, естественно, вырабатываются другие кодовые комбинации на выходе.

В случае отсутствия сигнала на входе устройства (например, во время пауз) изображающая точка должна находиться в начале координат. Логично предполагать, что кодирующее

устройство в этом случае должно выдавать нулевую кодовую комбинацию. Однако действие шумов приводит к тому, что в действительности точка отклоняется от начала координат и хаотически блуждает через все секторы. При этом кодирующее устройство выдает на выход соответствующие кодовые комбинации, что очень сильно искажает динамику последовательности кодовых комбинаций в целом. Подобное явление наблюдается также в случае измерения квазитопологических характеристик и времени пребывания. Однако степень влияния на окончательные

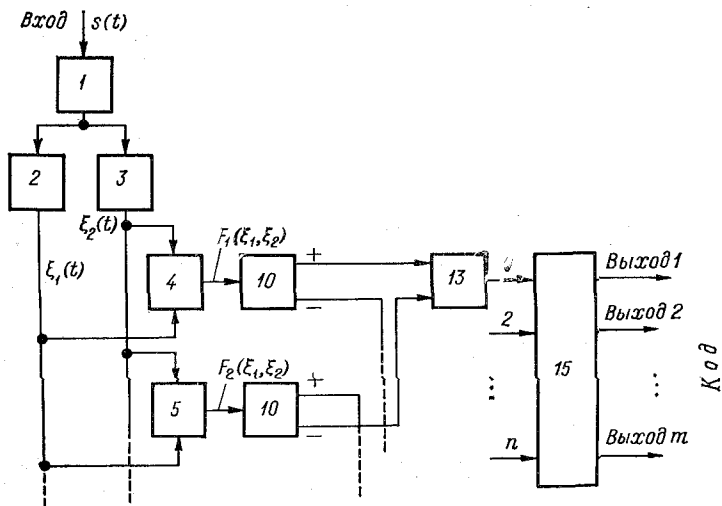


Рис. 22. Структурная схема устройства для кодирования фазовых изображений

результаты распознавания сигналов здесь меньше, так как квазитопологические характеристики и время пребывания являются более «грубыми» признаками, чем последовательности кодовых комбинаций. Поэтому при кодировании целесообразно ввести центральный сектор, определяемый только модулем радиус-вектора изображающей точки независимо от его угла поворота. Если выбрать диаметр этого сектора достаточно большим (чтобы превысить уровень шумов), то можно частично избавиться от искажений последовательностей кодовых комбинаций.

При использовании принципа построения схемы устройства, показанной на рис. 22, приходится формировать условия существования меандровых функций для каждого сектора в отдельности (так как это делается при измерении времени пребывания). Очевидно, при этом кодирующее устройство должно иметь столько входов, сколько имеется секторов. В целом устройство можно заметно упростить (уменьшить количество необходимых элементов для его построения), определенным образом изменяя

принцип образования кодов отдельных секторов. Для этого сначала вырабатываются кодовые комбинации для определенных полуплоскостей, образующихся при проведении граничных линий секторов. Кодовые комбинации отдельных секторов получаются при логическом суммировании комбинаций соответствующих полуплоскостей. Поясним этот принцип на примере кодирования трехразрядным двоичным кодом фазового изображения $\xi_1(t)$, $\xi_2(t)$, разделенного на 6 секторов тремя прямыми граничными линиями (рис. 23). Седьмой — центральный сектор — введен по причинам, упомянутым выше. (Отметим, что трехразрядный код допускает наличие восьми секторов, так что в нашем случае остается неиспользованной возможность введения еще одного сектора.)

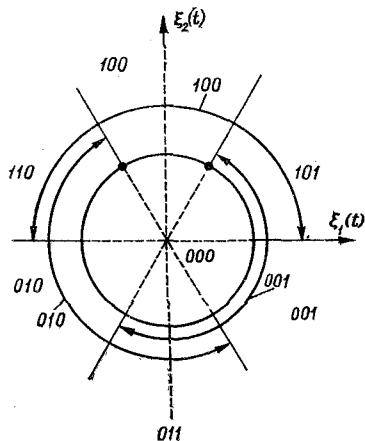


Рис. 23. К образованию кодовых комбинаций методом суммирования

Центральному сектору присваивается нулевая кодовая комбинация (000), а полуплоскостям, образованным при помощи граничных линий, — комбинации 001, 010, 100 соответственно. Наглядно видно, что комбинации отдельных секторов получаются при логическом суммировании комбинаций соответствующих полуплоскостей. Отметим, что при таком кодировании расстояние по Хеммингу между кодовыми комбинациями соприкасающихся секторов равно 1, а противоположных — 3 (за исключением центрального сектора).

Как было упомянуто в § 9, этим достигается большая помехоустойчивость.

Запишем условия пребывания изображающей точки в центральном секторе и в полуплоскостях:

$$\left. \begin{aligned} \text{в центральном секторе (код 000)} & - \xi_1^2(t) + \xi_2^2(t) - M < 0; \\ \text{в первой полуплоскости (код 001)} & - \xi_2(t) - k\xi_1(t) < 0; \\ \text{во второй полуплоскости (код 010)} & - \xi_2(t) + k\xi_1(t) < 0; \\ \text{в третьей полуплоскости (код 100)} & - \xi_2(t) > 0. \end{aligned} \right\} (22)$$

Здесь M — радиус центрального сектора, определяемый уровнем шумов.

Соответствующая этим условиям структурная схема устройства изображена на рис. 24. При помощи предварительного усилителя 1, функциональных преобразователей 2, 3, квадраторов 16 и сумматоров 8, 9, формируются функции $\xi_2(t)$, $\xi_2(t) - k\xi_1(t)$, $\xi_2(t) + k\xi_1(t)$, $\xi_1^2(t) + \xi_2^2(t) - M$, входящие в усло-

вия (22). Клиппирующие эти функции ограничители 10 настроены по-разному. В случае выполнения условия $\xi_1^2(t) + \xi_2^2(t) - M < 0$ на выходе соответствующего ограничителя устанавливается низкий (нулевой) уровень напряжения. В случае же выполнения остальных условий из (22) на выходах соответствующих ограни-

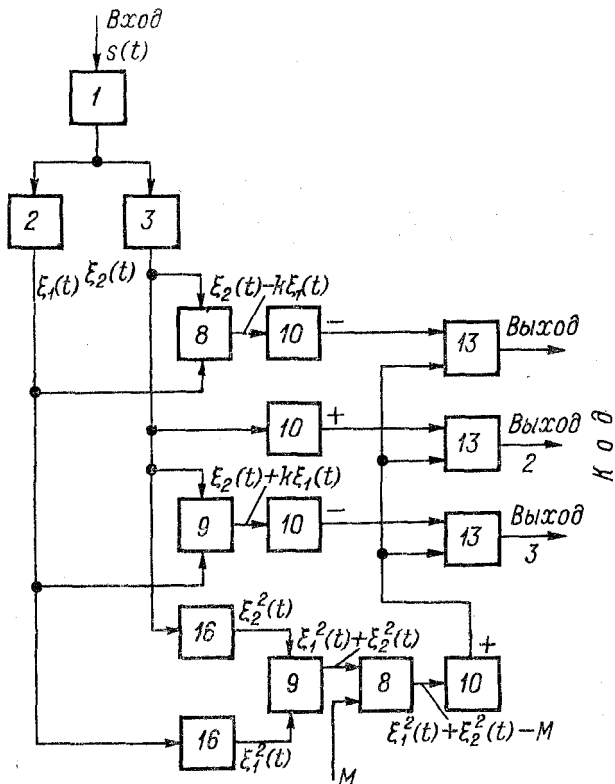


Рис. 24. Структурная схема устройства для кодирования фазовых изображений методом суммирования

чителей устанавливаются высокие (единичные) уровни напряжения. Нетрудно видеть, что при этом на выходах схем совпадения (схем И) 13 автоматически устанавливаются необходимые кодовые комбинации согласно рис. 23 и условиям (22).

При использовании принципа кодирования каждого сектора в отдельности пришлось бы моделировать шесть условий пребывания точки в секторах (вместо трех в рассмотренном случае) и, кроме того, строить кодирующее устройство (в рассмотренном случае необходимость в нем отпала).

15. Определение интегральных оценок

Рассмотрим способы выделения признаков, основанных на использовании интегральных оценок структурных связей с явным временем.

Очевидно, получение оценок (16) или (17) в чистом виде возможно цифровыми способами при помощи ЭВМ или при помощи интегратора на аналоговой основе. Однако проще всего реализуется не чистое интегрирование, а интегрирование с весом на основании использования связи между реакцией и возмущением некоторой линейной системы. Воспользуемся этой связью для составления алгоритма аппаратурного определения рассматриваемых оценок. Для этого выражение (16) перепишем в виде:

$$J_{\xi} = \int_t^{t+T} \xi(\tau) d\tau \Big|_{t=t_0}.$$

Здесь мы имеем дело со «скользящим» интегрированием, которое можно видоизменить введением «скользящей» весовой функции:

$$J_{\xi} = \int_0^t h(t-\tau) \xi(\tau) d\tau \Big|_{t=t_0},$$

где $h(t) = \sigma(t) - \sigma(t-T)$, а единичная функция

$$\sigma(t) = \begin{cases} 0, & t < 0; \\ 1, & t \geq 0. \end{cases}$$

Однако весовая функция в таком виде (в виде прямоугольного окна) физически неосуществима. Поэтому целесообразно расширить понятие интегральной оценки введением весовой функции $h(t)$ в виде импульсной характеристики физической осуществимой системы. (Тогда правомерно интегральную оценку назвать обобщенной.) В качестве такой системы может служить фактор низких частот, параметры которого определяются необходимым временем наблюдения сигнала T .

Аналогично дело обстоит и с интегральной оценкой вида (17).

Таким образом, обобщенные интегральные оценки аппаратурно определяются путем измерения значений реакции фильтра низких частот в определенные моменты времени при воздействии на его входе возмущений в виде функций $\xi(t)$ или $\mu(t)\xi(t)$. Измерение же значений реакции фильтра низких частот можно осуществить цифровым способом при помощи преобразователя аналог — код.

16. Определение некоторых структурных связей с явным временем

В этом параграфе рассмотрим вопросы, связанные с аппаратурным определением введенных в § 3 некоторых структурных связей с явным временем: диссипанта, консерванта, кондисанта, кондисквadrата.

В связи с разработкой перемножителей и делителей, действующих в четырех квадрантах и в широком диапазоне частот [23], возможно непосредственное моделирование соотношений (2). Соответствующие устройства, как видно из (2), состояли бы из дифференциаторов, перемножителей и делителей.

Можно также синтезировать схемы устройств согласно выражениям (3) [10]. Так, например, структурные схемы устройств

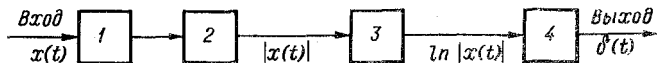


Рис. 25. Структурная схема устройства выделения диссипанта

для определения диссипанта и кондисанта приведены на рис. 25 и 26 соответственно. На этих рисунках 1 — предварительный усилитель, 2 — детектор для определения модуля или квадрата соответствующей функции, 3 — логарифматор, 4 — дифференциатор. Как видно из (2) и (3), консервант и кондисквадрат

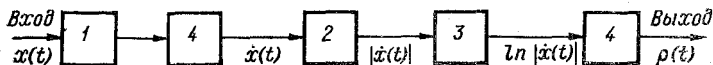


Рис. 26. Структурная схема устройства выделения кондисанта

без перемножителей и делителей выделить трудно. Можно пользоваться их логарифмическими функциями, получаемыми путем логарифмирования модулей диссипанта и кондисанта с последующим их сложением или вычитанием.

Глава четвертая

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ЭФФЕКТИВНОСТИ ПРИЗНАКОВ

17. Аппаратура и общая методика проведения экспериментов

С целью экспериментальной проверки теоретических предположений и выводов об эффективности использования структурных свойств речевых сигналов для их распознавания был проделан целый ряд экспериментов. Сущность этих экспериментов заключалась в аппаратном выделении разных систем признаков, описанных в предшествующих главах, с последующим моделированием распознавания ограниченного набора отдельно произнесенных слов-команд с помощью ЭВМ. Выбор слов в качестве основной единицы распознавания отражает стремление

авторов кратчайшим путем прийти к созданию автоматических «распознавателей» для практических применений.

Исследование эффективности признаков для распознавательных целей почти невысказимо без применения ЭВМ. Однако решение задачи целиком и полностью с помощью ЭВМ (начиная с выделения признаков и кончая классификацией) также не рационально в связи с большими затратами машинного времени. Особенно ярко эта проблема возникает в случае наличия ЭВМ средних возможностей (например, типа «Минск-22») и при необходимости обработки большого по объему речевого материала. Нам представляется, что наиболее подходящим решением в указанной ситуации является следующая последовательность действий: 1) выделение признаков с помощью специализированных устройств, 2) автоматический ввод выделенных признаков в ЭВМ, 3) классификация образов с помощью ЭВМ.

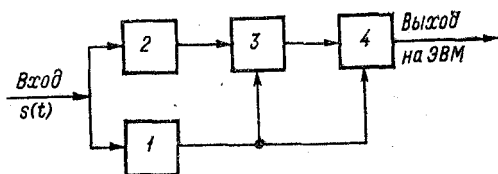


Рис. 27. Структурная схема устройства ввода признаков в ЭВМ

Структурная схема устройства, осуществляющего выделение и автоматический ввод признаков в ЭВМ, приведена на рис. 27. Речевой сигнал $s(t)$ из микрофона или магнитофона поступает на блок 2, представляющий собой одно из устройств выделения признаков, рассмотренных в третьей главе. В случае использования признаков, основанных на подсчете чисел пересечений или времени пребывания, блок 3 является блоком обыкновенных двоичных счетчиков, осуществляющих этот подсчет. В случае же использования кодов фазовых изображений или интегральных оценок структурных связей с явным временем, блок 3 — это блок буферных регистров памяти для запоминания на определенное время последовательностей кодовых комбинаций или интегральных оценок, преобразованных в цифровую форму; 4 — устройство согласования и управления режимом ввода признаков в ЭВМ; 1 — устройство, вырабатывающее метки начала и конца отрезков наблюдения сигнала (в частности, начала и конца слова) и управляющее работой счетчиков или буферной памяти, а также устройства согласования с ЭВМ.

В описанных в этой главе экспериментах использовалась машина «Минск-22» и признаки с помощью устройства ввода вводились прямо в оперативную память машины за исключением некоторых случаев, отмеченных отдельно.

Основной речевой материал состоял из отдельно произнесенных слов-команд. Всего использовалось 20 слов, включая названия цифр от нуля до девяти и 10 команд типа: «логарифми-

рование», «сложить», «вычесть», «стоп», «пуск» и др. Эксперименты в основном проводились с двумя группами слов: первая из них состояла только из названий цифр (10 различных слов), вторая включала также и остальные 10 команд. Слова произносились по несколько раз дикторами обоих полов.

В составлении речевого материала первой группы участвовало 30 мужчин и 10 женщин — всего 40 дикторов. Каждое слово произносилось ими по 3 раза. Таким образом общее число реализаций слов равнялось 1200. Произношения 23 мужчин и 7 женщин (всего 90 реализаций каждого слова) составили обучающую выборку, а произношения остальных 10 дикторов (всего 30 реализаций каждого слова) — контрольную выборку. Вторая группа состояла всего из 2000 реализаций, 20 слов, произнесенных 10 дикторами. Каждый диктор каждое слово произносил по 10 раз. Все 2000 реализаций использовались как для обучения, так и для контроля.

Предварительно слова записывались на магнитофонную ленту типа 2 при помощи магнитофона МАГ-59 и микрофона МД-47. Запись велась в тихом рабочем помещении, однако никакие специальные меры против акустических шумов не применялись. Никакой предварительный отбор дикторов по качеству произнесения или по каким-нибудь другим критериям не проводился и никакого предварительного обучения они не проходили. Единственное, что от них требовалось, — это произносить слова нормально, в привычной для себя манере и делать паузы 1—2 с между отдельными произнесениями.

Во время экспериментов слова воспроизводились с помощью того же магнитофона, что и при записи. После выделения и ввода признаков в ЭВМ проводилось обучение классификатора и классификация слов.

Классификация проводилась в основном на основе использования двух видов дискриминантных функций: линейных и квадратичных. Первые из них, приводящие к классификации по минимуму евклидова расстояния, имели вид [20]:

$$g_i(\mathbf{X}) = \mathbf{X} \cdot \mathbf{M}_i - \frac{1}{2} \mathbf{M}_i \cdot \mathbf{M}_i, \quad i = 1, 2, \dots, R.$$

Здесь \mathbf{X} — вектор-столбец в пространстве используемых признаков, представляющий собой классифицируемую точку (в нашем случае реализацию какого-нибудь слова); векторы $\mathbf{M}_1, \dots, \mathbf{M}_R$ описывают точки-эталоны, в качестве которых использовались только векторы средних классов, определяемые по обучающей выборке; R — число классов (в нашем случае — число различных слов).

Выбор классификатора по минимуму евклидова расстояния основывается на простоте его реализации как программным, так и аппаратным путем. Однако недостаточная мощность линейных дискриминантных функций привела к необходимости использования более мощных, квадратичных дискриминантных функций. Последние имели вид [1, 20]:

$$g_i(\mathbf{X}) = -(\mathbf{X} - \mathbf{M}_i)^t \sum_i^{-1} (\mathbf{X} - \mathbf{M}_i), \quad i = 1, 2, \dots, R,$$

где $(X - M_i)^t$ есть транспонированный вектор $(X - M_i)$ (вектор-строка); Σ_i^{-1} — матрица, обратная ковариантной матрице Σ_i (средние классов M_i и ковариантные матрицы Σ_i определяются по обучающей выборке).

Отметим, что подобные дискриминантные функции соответствуют классификатору по минимуму расстояния махалонобисского типа [1].

18. Исследование эффективности квазитопологических характеристик

Основная часть экспериментов проводилась на базе использования элементарного фазового изображения типа $\xi_1(t) = s(t)$, $\xi_2(t) = \dot{s}(t)$, пересекаемого восемью лучами связи, как показано на рис. 7. Практически применялось только семь лучей, так как числа пересечений третьего и седьмого лучей совпадают между собой (см. § 7). Схема соответствующего устройства выделения признаков показана на рис. 19. Время наблюдения сигнала устанавливалось равным длительности воспроизводимых слов. При этом блок I в устройстве ввода признаков (рис. 27) выполнял функции детектора начала и конца слова. Таким образом, каждое воспроизводимое слово представлялось семимерным вектором $X(n_1, \dots, n_6, n_8)$ в пространстве признаков. (Здесь компоненты вектора n_j , $j=1, \dots, 6, 8$ — это числа пересечений фазовым изображением соответствующих лучей.) Кроме эффективности абсолютных значений чисел пересечений была исследована также эффективность относительных их значений γ_j , определяемых по формуле $\gamma_j = n_j/T$, где T — длительность наблюдения сигнала, т. е. длительность реализации слова. (Методика измерения длительности реализаций слов T описана в следующем параграфе.)

Результаты моделирования распознаваний первой группы слов (названий цифр) приведены в табл. 1.

Таблица 1

Средняя эффективность распознавания 10 слов по точкам пересечений

Признаки	Дискриминантные функции	Эффективность, %	
		Обучающая выборка	Контрольная выборка
n_j	Линейные	76,9	69,3
n_j	Квадратичные	99,0	94,3
γ_j	Линейные	75,1	69,0
γ_j	Квадратичные	98,0	93,0

Табл. 2 иллюстрирует распознаваемость каждого слова в отдельности в случае использования квадратичных дискриминантных функций и признаков типа n_j . (Как видно из табл. 1, при этом достигнута наибольшая распознаваемость всех слов в целом.)

Средняя эффективность распознавания — слова названия
цифры по точкам пересечений

Слово	0	1	2	3	4
Эффективность при обучающей выборке, %	97,7	97,7	98,8	98,8	100
Эффективность при контрольной выборке, %	90,1	90,1	93,3	93,3	100

Продолжение

Слово	5	6	7	8	9
Эффективность при обучающей выборке, %	98,8	100	100	100	98,8
Эффективность при контрольной выборке, %	90,1	100	100	93,3	93,3

В табл. 3 приведены результаты распознавания второй группы слов (10 названий цифр и 10 команд).

Таблица 3

Средняя эффективность распознавания 20 слов
по точкам пересечений

Признаки	Дискриминантные функции	Эффективность, %
n_j	Линейные	65,8
n_j	Квадратичные	92,0
γ_j	Линейные	62,3
γ_j	Квадратичные	90,0

Анализ результатов указывает на целесообразность использования квадратичных дискриминантных функций, несмотря на сравнительную сложность их реализации как программным, так и аппаратным путем. Сравнение эффективности абсолютных и относительных значений чисел пересечений (n_j , γ_j) не дает преимуществ ни одним из них. Разница в надежности распознавания статистически не обоснована. Это наводит на мысль, что более выгодно пользоваться ненормированными признаками (n_j), так как при этом отпадает необходимость операции деления на T .

Одной из важнейших и необходимых предпосылок успешного решения задачи распознавания в целом является правильный выбор наиболее информативных признаков, т. е. таких признаков, в которых сосредоточена наиболее существенная для распознавания информации. В связи с этим на основании критерия средней дивергенции [12] было проделано упорядочение признаков n_j по мере уменьшения их информативности. Оказалось, что согласно принятому критерию признаки по своей информативности располагаются в следующий ряд: $n_5, n_1, n_3, n_6, n_2, n_4, n_8$. Наиболее наглядно вклад каждого признака в распознаваемость представлен данными табл. 4.

Таблица 4

Средняя эффективность распознавания 10 слов по точкам пересечений при различных сочетаниях признаков (дискриминантные функции квадратичные)

Признаки	Эффективность при контрольной выборке, %
$n_1, n_2, n_3, n_4, n_5, n_6, n_8$	94,3
$n_1, n_2, n_3, n_4, n_5, n_6$ —	94,0
n_1, n_2, n_3 — n_5, n_6 —	93,3
n_1 — n_3 — n_5, n_6 —	92,1
$n_1, — n_3 — n_5 — —$	88,8
$n_1 — — — n_5 — —$	75,6
— — — — $n_5 — —$	67,0

Приведенные данные свидетельствуют о том, что наиболее информативными являются признаки n_1, n_3, n_5 . Это соответствует числам пересечений полуосей абсцисс и ординат фазовым изображением (см. рис. 7). В какой-то мере это согласуется с данными визуального наблюдения за фазовыми изображениями. Самые характерные изменения изображений происходят именно в тех областях, которые пересекаются полуосями абсцисс и ординат.

Эффективность квазитопологических характеристик проверялась и в случае более сложных фазовых изображений, а именно: $s(t), \dot{s}(t)/s(t); s(t), \ddot{s}(t)/s(t); s(t), \dot{s}(t)/\dot{s}(t); \dot{s}(t)/s(t), \ddot{s}(t)/s(t)$. В качестве обучающей выборки (она же служила и для контроля) использовалось по 14 реализаций (произнесения 4 мужчин и 3 женщин) каждого слова речевого материала первой группы (всего 140 реализаций). Результаты моделирования распознавания примерно одинаковы для всех использованных фазовых изображений. В качестве иллюстрации в табл. 5 приводятся данные моделирования распознавания на базе использования изображения $\xi_1(t) = s(t), \xi_2(t) = \dot{s}(t)/s(t)$. Для классификации применялись квадратичные дискриминантные функции, а для упорядочения признаков — критерий средней дивергенции, как и в прежних экспериментах.

Количество правильно распознанных реализаций
10 слов по точкам пересечений на плоскости $s, s/s$
при различных сочетаниях признаков

Признаки	Количество правильно распознанных реализаций
$n_1, n_2, n_3, n_4, n_5, n_6, n_8$	138
$n_1, n_2, n_3, n_4, n_5, n_6$ —	138
n_1, n_2, n_3 — n_5, n_6 —	138
n_1 — n_3 — n_5, n_6 —	135
n_1 — n_3 — n_5 — —	130
n_1 — — — n_5 — —	119
— — — — n_5 — —	95

В табл. 6 даны результаты моделирования распознавания для каждого слова в случае использования сочетания признаков n_1, n_5 .

Таблица 6

Количество правильно распознанных реализаций каждого слова по точкам пересечения на плоскости $s, s/s$

Слово	0	1	2	3	4	5	6	7	8	9
Количество правильно распознанных реализаций	12	14	14	14	11	9	13	11	10	11

Результаты свидетельствуют о том, что в рамках приведенных экспериментов нельзя однозначно установить характерные особенности, а также преимущества одних фазовых изображений перед другими по информативности получаемых признаков. Очевидно для этого необходимы более тонкие эксперименты. В связи с относительной простотой реализации устройств выделения признаков при распознавании небольшого набора слов с эффективностью около 90% предпочтение все же следует отдавать элементарному фазовому изображению s, s .

19. Исследование эффективности признаков, основанных на подсчете времени пребывания

На основании использования этих признаков были проделаны эксперименты, аналогичные описанному в предыдущем параграфе. Выбранные для исследования фазовые плоскости были разделены на 8 равных секторов четырьмя граничными линиями

(см. рис. 13), и в каждом из них подсчитывалось время пребывания изображающей точки соответствующего фазового изображения в течение времени, равного длительности слова. (Соответствующие схемы устройств приведены и описаны в § 13.) Таким образом, каждое слово описывалось восьмимерным вектором $X(z_1, \dots, z_8)$ в пространстве признаков. (Здесь компоненты вектора $z_j, j=1, \dots, 8$ равны численным значениям времен пребывания в соответствующих секторах.) Исследовалась также эффективность относительных значений времени пребывания, определяемых соотношением $h_j = z_j/T$, где длительность реализации слова $T = \sum z_j$. Последнее выражение является непосредственным алгоритмом для определения длительности T . Результаты распознавания первой и второй группы слов приведены в табл. 7 и 8 соответственно.

Таблица 7

Средняя эффективность распознавания 10 слов по времени пребывания

Признаки	Дискриминантные функции	Эффективность, %	
		Обучающая выборка	Контрольная выборка
z_j	Линейные	71,3	64,0
z_j	Квадратичные	98,0	92,0
h_j	Линейные	73,2	66,5
h_j	Квадратичные	95,3	88,8

Таблица 8

Средняя эффективность распознавания 20 слов по времени пребывания

Признаки	Дискриминантные функции	Эффективность, %
z_j	Линейные	62,3
z_j	Квадратичные	92,6
h_j	Линейные	64,5
h_j	Квадратичные	90,0

Анализ результатов показывает, что, как и в случае квазитопологических характеристик, предпочтение следует отдавать квадратичным дискриминантным функциям и ненормированным признакам (z_j).

Были проведены исследования информативности признаков z_j по критерию энтропии распределений плотности вероятности образов в выборочном пространстве [12]. При этом принималось уменьшающее объем вычислений предположение о независимости признаков.

Оказалось, что наиболее информативными являются признаки z_1, z_4, z_5, z_6 , которые соответствуют времени пребывания в секторах, ограниченных осью абсцисс и линиями, проведенными к ней под углом 45° (см. рис. 13). Эффективность распознавания 20 слов только по признакам z_1, z_4, z_5, z_6 равнялась 55,6% при линейных дискриминантных функциях и 85,6% при квадратичных дискриминантных функциях.

На речевом материале, состоящем из 14 реализаций каждого слова названия цифры (всего 140 реализаций), проверялась эффективность признаков z_j ; также для плоскостей $s, \dot{s}/s; s, \ddot{s}/s; s, \dot{s}/\dot{s}; \dot{s}/s, \ddot{s}/s$. Результаты сведены в табл. 9.

Таблица 9

Количество правильно распознанных реализаций 10 слов по времени пребывания на плоскостях $s, \dot{s}/s; s, \ddot{s}/s; s, \dot{s}/\dot{s}; \dot{s}/s, \ddot{s}/s$ (дискриминантные функции квадратичные)

Плоскость	Количество правильно распознанных реализаций
$s, \dot{s}/s$	135
$s, \ddot{s}/s$	138
$s, \dot{s}/\dot{s}$	138
$\dot{s}/s, \ddot{s}/s$	139

Результаты экспериментов свидетельствуют о принципиальной возможности использования времени пребывания в секторах различных фазовых плоскостей в качестве признаков для распознавания небольшого набора слов. Очевидно для более тонкого анализа свойств отдельных плоскостей требуются дальнейшие эксперименты.

20. Исследование эффективности интегральных оценок

Рассматривалась эффективность интегральных оценок структурных связей с явным временем: диссипанта $\delta_s(t) = \dot{s}(t)/s(t)$, консерванта $\kappa_s(t) = \ddot{s}(t)/s(t)$, кондисанта $\rho_s(t) = \ddot{s}(t)/\dot{s}(t)$ и кондисквadrата $\lambda_s(t) = \ddot{s}(t)s(t)/\dot{s}^2(t)$.

Речевой материал, на котором исследовалась эффективность, состоял из 140 реализаций 10 слов названий цифр (по 14 реализаций каждого слова), которые использовались и для обучения, и для контроля. Воспроизводимые магнитофоном слова при помощи преобразователя аналог — код вводились в оперативную память ЭВМ, где в первую очередь осуществлялось формирование перечисленных структурных связей. Далее каждая реализация структурной связи разбивалась на 8 равных отрезков и в каждом из них подсчитывалась соответствующая оценка согласно выражению (16) (число отрезков выбрано эмпирически). Таким образом, каждая реализация любого слова представлялась восьмерным вектором $\mathbf{X} (r_1^{(\xi)}, \dots, r_8^{(\xi)})$, где $r_i^{(\xi)}$,

$j=1, \dots, 8$ — интегральные оценки соответствующих структурных связей $\xi(t)$ на отрезках j .

В эксперименте использовались квадратичные дискриминантные функции. Результаты сведены в табл. 10.

Таблица 10

Количество правильно распознанных реализаций 10 слов по интегральным оценкам структурных связей $\delta_s(t), \kappa_s(t), \rho_s(t), \lambda_s(t)$

Структурная связь	Количество правильно распознанных реализаций
$\delta_s(t)$	130
$\kappa_s(t)$	128
$\rho_s(t)$	131
$\lambda_s(t)$	133

Полученная эффективность распознавания по интегральным оценкам сравнима с эффективностью распознавания по другим рассмотренным системам признаков. Несомненна принципиальная возможность использования структурных связей с явным временем для выделения признаков при распознавании.

21. Исследование эффективности смешанных признаков

Результаты исследований различных систем признаков на информативность наводят на мысль о целесообразности объединения наиболее информативных признаков в одну систему с целью повышения эффективности распознавания. В связи со сравнительной простотой аппаратной реализации устройств выделения признаков эксперименты проводились на основании объединенной системы признаков $n_1, n_3, n_5, z_1, z_4, z_5, z_8$ (см. параграфы 18 и 19), в дальнейшем называемой смешанными признаками. Использовались квадратичные дискриминантные функции. Отличительной чертой этой серии экспериментов было то, что дикторы в микрофон говорили прямо в машинном зале. В составлении обучающей выборки участвовало 10 дикторов, и каждый из них каждое из 20 слов произносил по 10 раз. Программа контроля была организована так, что сразу после произнесения слова машина печатала результат распознавания. Это приблизило условия проведения эксперимента к предполагаемым реальным условиям эксплуатации распознающих систем. Средняя эффективность распознавания 2000 контрольных реализаций 20 слов составила 92%. Надо подчеркнуть, что при этом отношение сигнал-шум не превышало 20 дБ.

Специальные исследования показали, что данная система признаков особенно чувствительна к темпу произнесения. По-

этому логическим путем продолжения исследований следовало бы считать эксперименты по нормализации длительности реализаций слов путем автоматического исключения некоторых стационарных участков сигналов, как наиболее чувствительных к темпу произнесения на основании их предварительной сегментации (см. § 11).

ЗАКЛЮЧЕНИЕ

Проделанная работа, особенно ее экспериментальная часть, по мнению авторов, демонстрирует лишь принципиальную возможность использования структурных свойств речевых сигналов для их распознавания. Однако этим не ограничиваются потенциальные возможности структурных методов. Поэтому достигнутый этап исследований надо считать лишь начальным шагом как в экспериментальном, так и в теоретическом плане.

Во-первых, здесь следует отметить, что теория статистических решений не может быть единственной или общей основой решения проблемы распознавания речи. Учет только усредненных характеристик в процедурах обучения и распознавания речи неизбежно приводит к тому, что система также только в среднем справляется с задачами распознавания, допуская при этом индивидуальные ошибки. Более перспективными являются адаптивные системы распознавания речевых сигналов, которые так или иначе способны подстраиваться при изменяющейся ситуации. Таковой, например, является система, при которой удачно найденные признаки, обеспечивающие правильную классификацию, продолжают использоваться машиной, а неудачно найденные, приводящие к ошибкам, забываются. Особенно перспективна, на наш взгляд, текущая подстройка распознающей системы под диктора. Эта проблема представляет самостоятельный интерес.

В этом плане структурные методы таят в себе много неисчерпанных возможностей, в том числе сравнительную простоту и удобство выявления различных систем признаков сигнала, а также слежения за их динамикой и взаимодействием. Следует подчеркнуть также возможность получения систем признаков, инвариантных к различным факторам. Дальнейшие исследования должны быть направлены на более детальное изучение именно этих вопросов.

Авторы будут считать свою задачу выполненной, если эта книга будет поводом для критического пересмотра традиционных представлений о речевых сигналах и переключит внимание специалистов на более перспективные методы структурного анализа.

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. М., «Статистика», 1974. 240 с. с ил.
2. Андронов А. А., Витт А. А., Хайкин С. Э. Теория колебаний. М., Физматгиз, 1959. 916 с. с ил.
3. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., «Наука», 1965. 464 с. с ил.
4. Величко В. М., Загоруйко Н. Г. Распознавание больших словарей. — «Труды VIII Всесоюзной школы — семинара «Автоматическое распознавание слуховых образов». Ч. 3. Львов. Изд-во ЛГУ, 1974, с. 9—13 с ил.
5. Винцюк Т. К., Пучкова Н. Г. Распознавание слов речи с помощью обучаемого алгоритма. — «Труды VII Всесоюзной школы — семинара «Автоматическое распознавание слуховых образов», Алма-Ата, Изд-во АН Казахской ССР, 1973, с. 16—20 с ил.
6. Восприятие речи в распознающих моделях. Львов, Изд-во ЛГУ, 1971. 187 с. с ил. Авт.: М. Деркач, Р. Гумецкий, Л. Мишин и др.
7. Гребнева Н. И. Использование переходов между звуками для автоматического распознавания речи. — «Вопросы радиоэлектроники. Техника проводной связи», 1968, вып. 7, с. 9—21 с ил.
8. Гудонавичюс Р. В., Кемешис П. П., Читавичюс А. Б. Автоматическое распознавание слов в конечном наборе по их структурным связям с неявным временем. — «Акустический журнал», 1973, Т. 19, вып. 6, с. 824—829 с ил.
9. Гудонавичюс Р. В., Заездный А. М., Читавичюс А. Б. Распознавание слов на основе использования структурных связей с явным временем. — «Техническая кибернетика», 1973, № 2, с. 214—215.
10. Драугелис Л. П., Жвинис П. З. Аппаратурное выделение некоторых структурных функций с явным временем. — В кн.: Техническая кибернетика. Каунас, КПИ, 1973, с. 163—166 с ил.
11. Жуков С. Я., Жукова М. Г., Чистович Л. А. Некоторые новые представления относительно слухового анализа звукового потока. — «Акустический журнал», 1974, т. 20, вып. 3, с. 386—392 с ил.
12. Загоруйко Н. Г. Методы распознавания и их применение. М., «Советское радио», 1972. 206 с. с ил.
13. Заездный А. М., Плоткин Е. И., Черкасский Ю. А. Основы разделения и измерения сигналов по структурным свойствам. Л., ЛЭИС им. Бонч-Бруевича, 1971. 124 С. с ил.

14. **Заездный А. М.** Классификация сигналов, распространяющихся в водной среде, на основе их обработки по структурным свойствам.— В кн.: Труды четвертой всесоюзной школы—семинара по статистической гидроакустике. Новосибирск, Изд. института математики СО АН СССР, 1973, с. 14—22.
15. **Заездный А. М., Шейнман Е. И.** Законы и свойства распределений структурных связей случайных процессов.— В кн.: Тезисы докладов VI Всесоюзного симпозиума «Методы представления и аппаратный анализ случайных процессов и полей». Л., ВНИИЭП, 1973, с. 39—50 с ил.
16. **Куля В. И., Пирогов А. А.** О фонетическом коде речевого сигнала.— «Радиотехника», 1970, т. 25, № 6, с. 29—34 с ил.
17. **Лабукин В. К., Молчанов А. П.** Модели механизмов слуха. М., «Энергия», 1973, 200 с. с ил.
18. **Модель восприятия речи человеком.** Новосибирск, «Наука», 1969. 60 с. с ил. Авт.: Л. В. Бондарко, Н. Г. Загоруйко, В. А. Кожевников и др.
19. **Молчанов А. П., Томах В. Ф.** Исследование модели механизма сегментации речевых сигналов на фонемы.— «Труды VI Всесоюзной школы—семинара «Автоматическое распознавание слуховых образов». Таллин, Изд-во АН ЭССР, 1972, с. 136—144 с ил.
20. **Нильсон Н.** Обучающиеся машины. М., «Мир», 1967. 180 с. с ил.
21. **Руджионис А., Кемешис П. П.** Исследование синтеза речи по огибающим и фазовым функциям.— В кн.: Автоматика и вычислительная техника. Т. 5. Вильнюс, «Минтис», 1972, с. 63—70 с ил.
22. **Смирнов В. И.** Курс высшей математики. Т. 2. М., «Наука», 1974. 656 с. с ил.
23. **Современные методы и средства обработки сигналов.** Л., ЛЭИС им. М. А. Бонч-Бруевича, 1971. 147 с. с ил. Авт.: А. М. Заездный, Е. И. Плоткин, О. Н. Работкин и др.
24. **Стинрод Н., Чинн У.** Первые понятия топологии. М., «Мир», 1967. 224 с. с ил.
25. **Тихонов В. И.** Выбросы случайных процессов. М., «Наука», 1970, 392 с. с ил.
26. **Флананган Дж. Л.** Анализ, синтез и восприятие речи. М., «Связь», 1968. 396 с. с ил.
27. **Цемель Ф. И.** Опознавание речевых сигналов. М., «Наука», 1971. 148 с. с ил.

ОГЛАВЛЕНИЕ

Предисловие	3
Глава первая. Структурная модель речевого сигнала	5
1. Структурные свойства и структурные связи	—
2. Предварительные преобразования речевого сигнала	8
3. Некоторые структурные связи	9
4. Соображения об адекватности структурной модели речевого сигнала моделям механизмов слуха	12
Глава вторая. Признаки для распознавания	16
5. Общие положения	—
6. Топологические инварианты	18
7. Квазитопологические характеристики	21
8. Время пребывания изображающей точки фазового изображения в секторах плоскости	27
9. Кодирование фазовых изображений	32
10. Интегральные оценки структурных связей с явным временем	—
11. Некоторые вопросы предварительной сегментации речевых сигналов	34
Глава третья. Выделение и измерение признаков	36
12. Измерение квазитопологических характеристик	—
13. Измерение времени пребывания	40
14. Образование кодов фазовых изображений	43
15. Определение интегральных оценок	48
16. Определение некоторых структурных связей с явным временем	—
Глава четвертая. Экспериментальные исследования эффективности признаков	49
17. Аппаратура и общая методика проведения экспериментов	—
18. Исследование эффективности квазитопологических характеристик	52
19. Исследование эффективности признаков, основанных на подсчете времени пребывания	55
20. Исследование эффективности интегральных оценок	57
21. Исследование эффективности смешанных признаков	58
Заключение	59
Список литературы	60