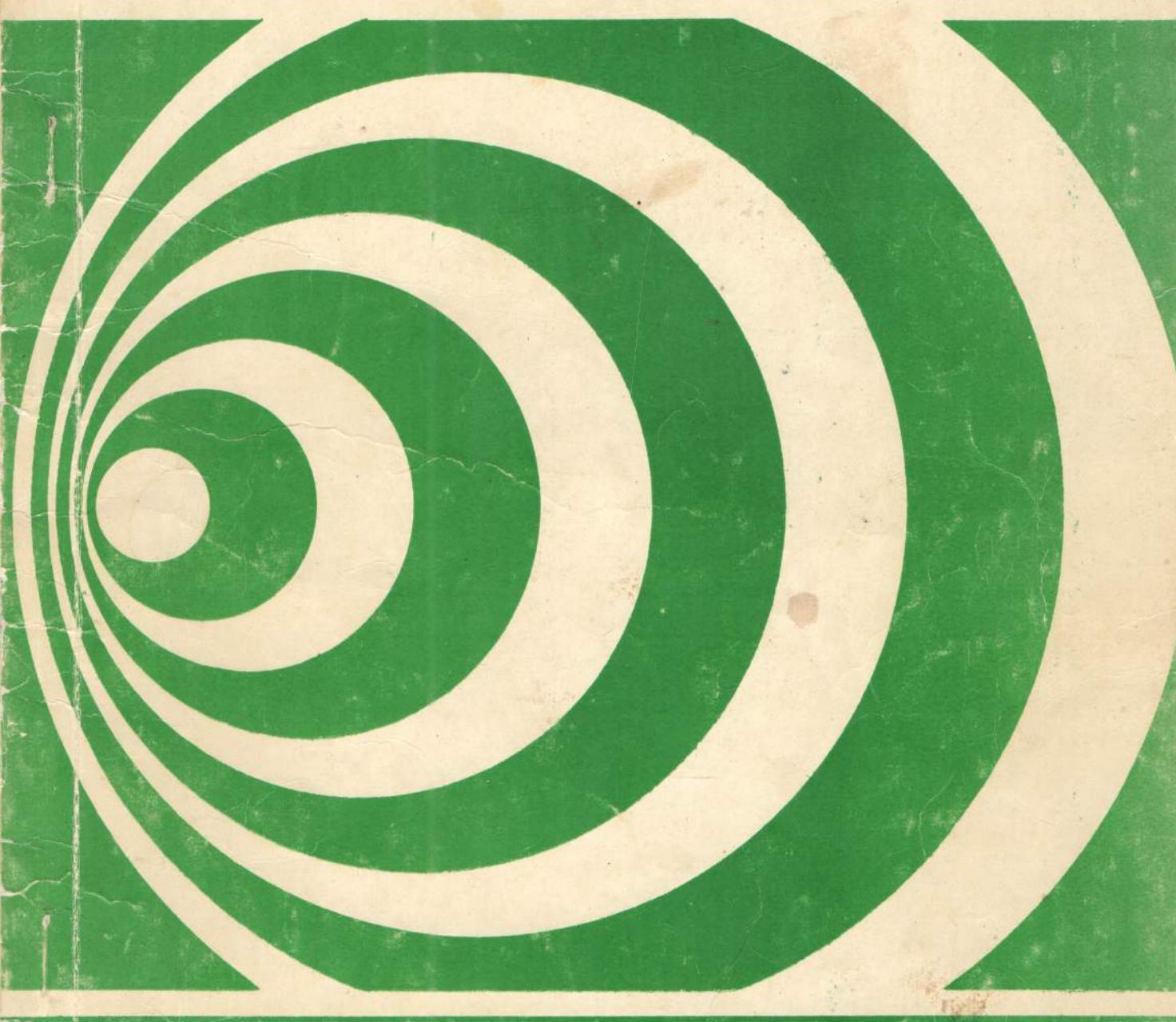


ГОСУДАРСТВЕННЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ РАДИО



ПРИЛОЖЕНИЯ
НИИ Р

№ 4

1969

ГОСУДАРСТВЕННЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ РАДИО

ТРУДЫ НИИР

ТРУДЫ НИИР

ВЫПУСК 4

1969

Ответственный редактор
доктор технических наук, профессор
ФОРТУШЕНКО А.Д.

Редакционная коллегия: канд.техн.наук АФАНАСЬЕВ Ю.А., докт.техн.
наук БОРОДИЧ С.В., канд.техн.наук БЫКОВ В.Л., докт.техн.наук ГУРЕВИЧ М.С.,
канд.техн.наук КАНТОР Л.Я., канд.техн.наук КОРОБКОВ Л.А., докт.техн.наук
КРИВОШЕЕВ М.И., докт.техн.наук КУЗНЕЦОВ В.Д., канд.экон.наук МАНАЕНКОВ
С.Д., докт.техн.наук МОДЕЛЬ А.М., канд.техн.наук НЕМИРОВСКИЙ А.С.,
докт.техн.наук ПИРОГОВ А.А., канд.техн.наук ЧЕРНОВ Ю.А., канд.техн.наук
ШАМШИН В.А., канд.техн.наук ШЛЮГЕР И.С.,

Постро-
венно связ-
ход распо-
времени

в дискрет-
та [2], [3]

Кажды

где t_o^k —
 $T^k = t_i^k$ —
Задача
радиопрежем
[4], [5]
произношен
Пусть этап

где τ
ства:

Государственный научно-исследовательский институт радио
Труды НИИР № 4, 1969 г.

Редакторы: Л.Н.БЫКОВА, А.З.БОРОВАЯ

Корректор А.Е.ШВАРЦ

Подписано в печать 26.П.1970 г. Т-02971 зак.№ 2

Тираж 700 экз. Цена 95 коп. 19,5 п.л.

где S
R

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ ЗВУКОСОЧЕТАНИЙ В ТЕКУЩЕМ РЕЧЕВОМ СИГНАЛЕ

Решается задача текущего распознавания звукосочетаний в непрерывном потоке речи как задача различения набора сигналов среди помех с неизвестным началом, длительностью и нелинейным масштабом времени. Приводится функциональная блок-схема автомата, реализующего описываемый математический аппарат принятия решения.

Построение фонемных вокодеров и систем, управляемых речью, непосредственно связано с задачей преобразования речевого сигнала, поступающего на вход распознающего устройства в виде пучка ρ непрерывно меняющихся во времени t нормализованных параметров [1]

$$\xi(t) = [\xi_1(t), \xi_2(t), \dots, \xi_p(t)], \quad (1)$$

в дискретную последовательность номеров звукосочетаний из конечного алфавита [2], [3]

$$\lambda_1, \lambda_2, \dots, \lambda_k, \dots \quad (1 \leq \lambda_k \leq M).$$

Каждый из этих номеров ставится в соответствие с отрезком сигнала (1):

$$\lambda_k \longleftrightarrow S^k = \{G^k, \xi(t)\}, \quad G^k = [t_o^k, t_1^k], \quad (2)$$

где t_o^k — начало λ_k -го звукосочетания в сигнале (1),

$T^k = t_1^k - t_o^k$ — длительность реализации звукосочетания.

Задача (2) может быть рассмотрена с точки зрения теории оптимального радиоприема как задача различения M сигналов при наличии помех и искажений [4], [5] с неизвестными λ, t_o и t_1 . Будем считать гауссовыми отклонения произношения дикторов от эталонного, остающиеся после нормализации [1]. Пусть эталоны звукосочетаний заданы в виде

$$\xi^k = \{G^k, \bar{\xi}^k(\tau)\}, \quad G^k = [0, T^k], \quad (3)$$

где T — эталонное время. Тогда задача (2) сводится к решению неравенства:

$$\min_{\lambda} h^k R^2(S^k, \xi^k) < N, \quad (4)$$

где S — искомые отрезки сигнала (1),

R — некоторый функционал, определяющий расстояние между сигналами,

λ, H - вес и порог, задаваемые на основе априорных сведений [6] и выбранного критерия оптимальности [4].

Вычисление расстояния R по обычной формуле среднеквадратичного отклонения сигналов [4], [5] невозможно, так как искомые отрезки (2) совпадают с эталонами (3) лишь с точностью до нелинейного изменения масштаба времени. Пользуясь几乎到处都是

$$C = \{t(c), \tau(c)\}, \quad \frac{dt}{dc} \geq 0, \quad \frac{d\tau}{dc} \geq 0, \quad (5)$$

отрезков (2) и (3) друг в друга с сохранением направления отсчета [7], определим расстояние R функционалом

$$R^2(S, \xi^\lambda) = \inf_C \left\{ \frac{1}{\mu C} \int_C \tau^2 [t(c), \tau(c)] d\mu \right\}, \quad (6)$$

где $\tau(t, \tau)$ - расстояние между отсчетами сигнала (1) и эталона (3) в моменты t и τ .

$d\mu = \max(dt, d\tau)$ - мера времени на отображении C .

В связи с отличием задачи текущего распознавания от задачи с известными границами [7], в ф-ле (6) величина τ носит несколько иной смысл. Рассмотрим подробнее метрическое пространство сигналов. В p -мерном векторном фазовом пространстве

$$\Phi = \{\bar{X}\}, \quad \bar{X} = (x_1, x_2, \dots, x_p) \quad (7)$$

определим фазовое расстояние ρ между векторами \bar{X}^1 и \bar{X}^2 формулой

$$\rho^2(\bar{X}^1, \bar{X}^2) = \sum_{i=1}^p \alpha_i (x_i^1 - x_i^2)^2, \quad (8)$$

где α_i - весовые коэффициенты. Назовем пространством отсчетов ($p+1$)-мерное произведение Ω времени t на фазовое пространство (7)

$$\Omega = t \times \Phi = \{\bar{\omega}\}, \quad \bar{\omega} = (t, \bar{X}) = (t, x_1, x_2, \dots, x_p). \quad (9)$$

Расстояние τ между отсчетами $\bar{\omega}^1$ и $\bar{\omega}^2$ пространства (9) определим по формуле

$$\tau^2 = (\bar{\omega}^1, \bar{\omega}^2) = \delta^2(t^1, t^2) + \rho^2(\bar{X}^1, \bar{X}^2), \quad (10)$$

где $\delta = \sqrt{\alpha_0 (t^1 - t^2)^2}$ - расстояние по времени,

ρ - фазовое расстояние (8) (рис.1).

Будем говорить, что задан конечный сигнал

$$S = \{G, \xi(t)\}, \quad G = [t_0, t_1], \quad (11)$$

и ум если указаны отрезок G времени его существования и определенная на этом отрезке ρ - мерная непрерывная вектор-функция времени $\xi(t)$. Нетрудно видеть, что сигнал (11) представляется в $(\rho + 1)$ - мерном пространстве отсчетов (9) в виде пространственной кривой, растянутой вдоль оси времени t и взаимно однозначно проектируемой на отрезок G (рис.2). Множество сигналов (11) образует функциональное пространство конечных сигналов, определенных на отрезках времени с произвольными началом и длительностью $T=t_2-t_1$. Будем говорить, что отсчет $\bar{\omega}=(t, \bar{x})$ принадлежит сигналу (11), $\bar{\omega} \in S$, если $t \in G$ и $\bar{x}=\xi(t)$. Для вырожденного случая, когда один из сигналов превращается в отсчет $\bar{\omega}^0$, примем за расстояние между ним и сигналом (11) среднеквадратичное расстояние (10) от $\bar{\omega}^0$ до всех $\bar{\omega} \in S$ (рис.3):

$$R^2(\bar{\omega}^0, S) = \frac{1}{T} \int_G \tau^2(\bar{\omega}^0, \bar{\omega}) dt.$$

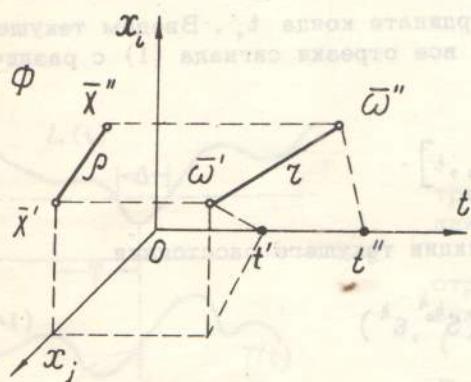


Рис.1

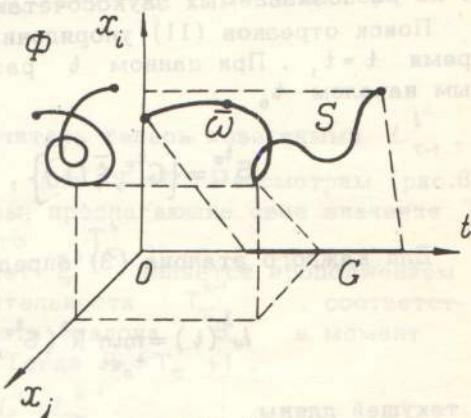


Рис.2

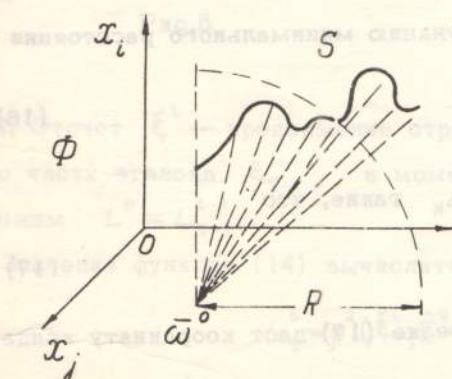


Рис.3

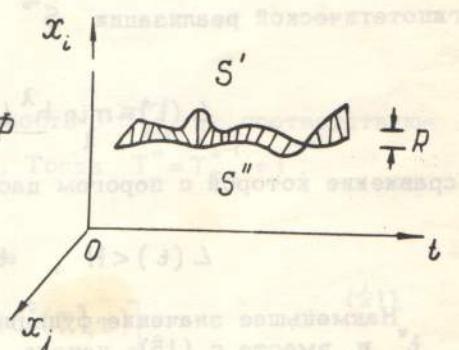


Рис.4

И, наконец, для двух сигналов S' и S'' будет (рис.4):

$$R^2(S', S'') = \inf_G \left\{ \frac{1}{\mu C} \int_G \tau^2(\bar{\omega}', \bar{\omega}'') d\mu \right\}, \quad (12)$$

Может
и учсть в
мости выч
Рассмо

где t — расстояние (10) между отсчетами $\bar{\omega} \in S'$ и $\bar{\omega}' \in S''$,
поставленными в соответствие отображением (5). Формулу (12) будем исполь-
зовать для определения расстояния (6) между эталоном (3) и отрезком (11)
сигнала (1), считая

$$T = t - t_0, \quad (13)$$

где t_0 — начало отрезка (11).
Решение задачи (4) состоит из четырех действий:
а) перебор всех отрезков (11) сигнала (1);
б) вычисление расстояний между выбранным отрезком (11) и всеми этало-
нами (3), (13);

в) выбор ближайшего эталона;
г) сравнение полученного минимального расстояния с порогом H для принятия решения о том, что отрезок (11) действительно является реализацией одно-
го из распознаваемых звукосочетаний.

Поиск отрезков (11) упорядочим по координате конца t . Введем текущее время $t = t_1$. При данном t рассмотрим все отрезки сигнала (1) с различным началом t_0 :

$$S^{t_0} = \{G^{t_0}, \bar{\xi}(t)\}, \quad G^{t_0} = [t_0, t].$$

Для каждого эталона (3) определим функцию текущего расстояния

$$L^\lambda(t) = \min_{t_0 \leq t} R^2(S^{t_0}, \varepsilon^\lambda) = R^2(S^{t_0}, \varepsilon^\lambda) \quad (14)$$

и текущей длины

$$T^\lambda(t) = t - t_0^\lambda \quad (15)$$

гипотетической реализации $S^{t_0^\lambda}$. Введем функцию минимального расстояния

$$L(t) = \min_\lambda h^\lambda L^\lambda(t), \quad (16)$$

сравнение которой с порогом даст отрезки Δ_k такие, что

$$L(t) < H, \quad t \in \Delta_k. \quad (17)$$

Наименьшее значение функции (16) на отрезке (17) даст координату конца t_i^κ и, вместе с (15), начала

$$t_0^\kappa = t_i^\kappa - T^{\lambda_m}(t_i^\kappa)$$

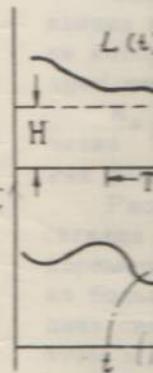
k -го отрезка сигнала (1), реализующего λ_m -е звукосочетание (рис.5).

Приведем рекуррентный алгоритм для вычисления функций (14), (15).
Пусть Δt — шаг дискретизации по времени. Этalon (3) примет вид матрицы

$$\varepsilon^\lambda = \{\bar{\omega}^\tau\} = \{(\tau, \bar{\xi}^\tau)\}, \quad \tau=0,1,\dots,n^\lambda = \frac{T^\lambda}{\Delta t}. \quad (18)$$

Будем одновременно в мости выч
одноименное
ным эталон
поступивши
вычисляют
Положим

Будем
одновремен
тельности
ным эталон
поступивши
вычисляют
Положим



3. Отсче
щего частк
Положим

Значени

Значени
(21) гипоте
На рис.
ки автомата
и 4 выхода.
на следующ
— со в

Множитель перед интегралом (6) можно опустить, так как $\mu C \approx T^\lambda$, и учесть в (16). Индекс λ в описании алгоритма опустим ввиду независимости вычисления функций (14), (15) для всех λ .

Рассмотрим $(n+1)$ частичных эталонов с общим началом в 0:

$$\mathcal{E}_\tau = \{\bar{\omega}^0, \dots, \bar{\omega}^\tau\}, \quad \tau = 0, 1, \dots, n. \quad (19)$$

Будем вычислять функции (14), (15) для всех частичных эталонов (19) одновременно. Пусть для момента $(t-1)$ текущего времени известны длительности $T_{\tau-1}^{t-1}$ отрезков сигнала (1), оптимально соответствующих частичным эталонам (19), и значения $L_{\tau-1}^{t-1}$ функции (14). Пусть ξ^t — вновь поступивший отсчет сигнала (1). Новые значения L_τ^t и T_τ^t для всех τ вычисляются, начиная с $\tau=0$, по имеющимся старым $L_{\tau-1}^{t-1}$ и $T_{\tau-1}^{t-1}$. Положим

$$T_0^t = 0, \quad L_0^t = \rho^2(\xi^t, \bar{E}^0). \quad (20)$$

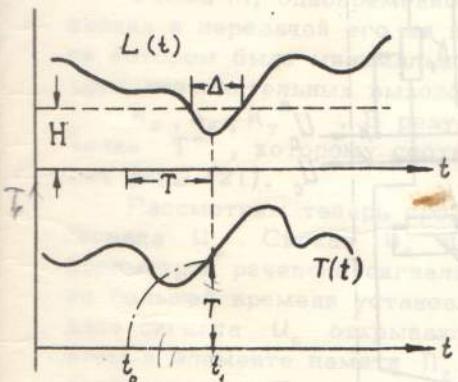


Рис.5

Будем считать теперь известными $L_{\tau-1}^t$, $T_{\tau-1}^t$ для $\tau > 0$. Рассмотрим (рис.6) три гипотезы, предлагающие свое значение T^k для искомого T_τ^t .

1. Отсчет ξ^t является продолжением отрезка длительности $T_{\tau-1}^{t-1}$, соответствующего части эталона $\mathcal{E}_{\tau-1}$ в момент $(t-1)$. Тогда $T^1 = T_{\tau-1}^{t-1} + 1$.

Положим $L^1 = L_{\tau-1}^{t-1}$.

2. Отрезок длительности $T_{\tau-1}^t$ соответствует частичному эталону \mathcal{E}_τ . Тогда $T^2 = T_{\tau-1}^t$. Положим $L^2 = L_{\tau-1}^t$.

3. Отсчет ξ^t — продолжение отрезка длительности $T_{\tau-1}^{t-1}$, соответствующего части эталона $\mathcal{E}_{\tau-1}$ в момент $(t-1)$. Тогда $T^3 = T_{\tau-1}^{t-1} + 1$. Положим $L^3 = L_{\tau-1}^{t-1}$.

Значение функции (14) вычисляется по формуле

$$L_\tau^t = \rho^2(\xi^t, \bar{E}^\tau) + \min_{K=I, II, III} [L^K + \delta^2(T^K, \tau)]. \quad (21)$$

Значение T_τ^t берется равным T^m , где m — номер выбранной в (21) гипотезы.

На рис.7 изображена функциональная схема стандартной элементарной ячейки автомата, реализующей вычисление L_τ^t и T_τ^t . Ячейка имеет 10 входов и 4 выхода. Все входы удобно разделить по типу источников входных сигналов на следующие 3 группы:

- со внешними источниками, общими для всех ячеек (ξ^t, U_c, U_c^*, U_o) ;

— с дополнительными источниками, общими для ячеек I и II ($T, T^1, T^2, T^3, L, L^1, L^2, L^3$);

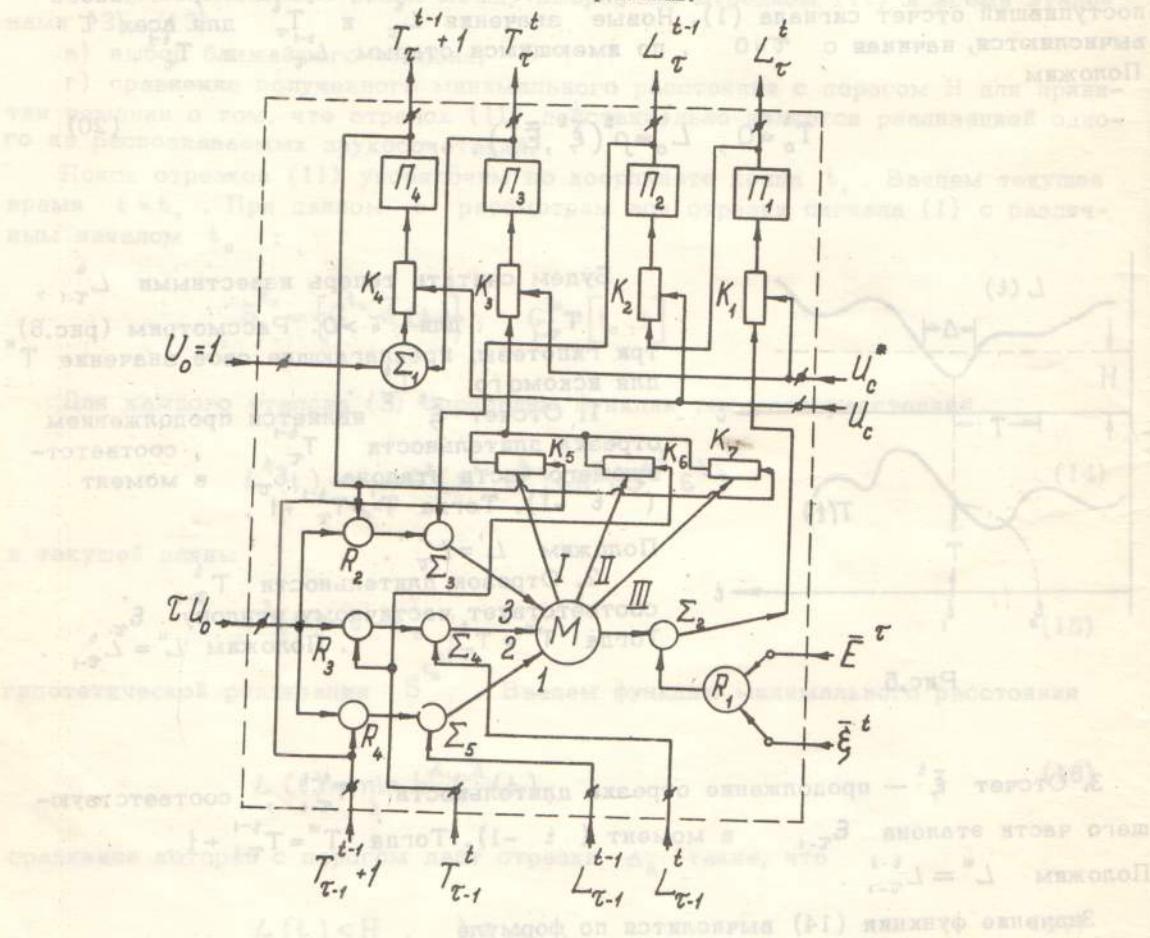
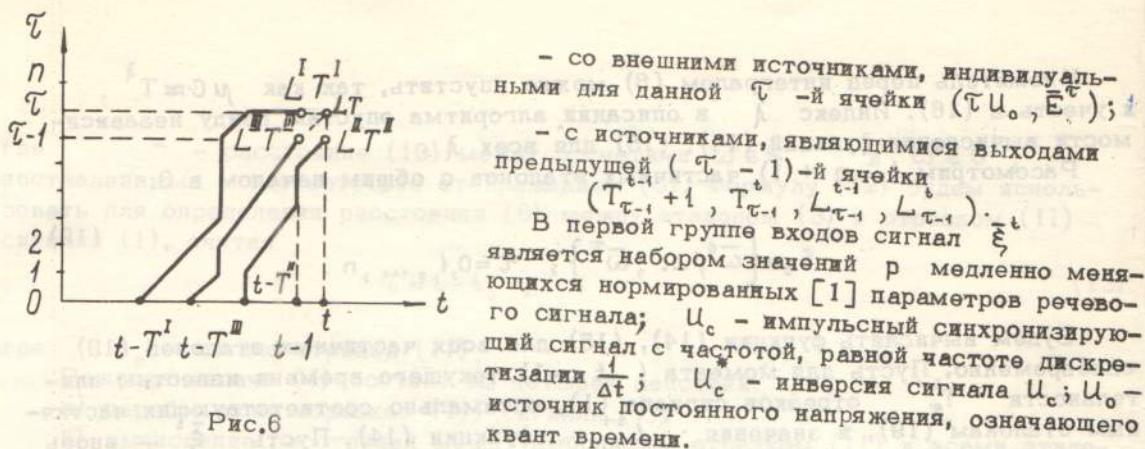


Рис.7

Во второй группе входов сигнал \bar{E}^τ является набором из p постоянных напряжений, величины которых совместно с сигналом τU_0 описывают τ -й отсчет эталона (3).

В третьей группе входов сигналы $T_{\tau-1}^{t-1}, T_{\tau-1}^t, L_{\tau-1}^{t-1}$,

$L_{\tau-1}^t$ являются выходными сигналами ($\tau - 1$)-й ячейки, изменяющими свои значения при появлении синхронизирующих импульсов U_c .

Все выходы ячейки удобно разделить на две группы: основные (L_{τ}^t, T_{τ}^t)

и вспомогательные ($L_{\tau}^{t-1}, T_{\tau}^{t-1}, T_{\tau}^{t-1} + 1$). Сигналы с основных выходов несут информацию о длительности T_{τ}^t отрезка сигнала (1), оптимально соответствующего τ -му частичному эталону (19), и о значении L_{τ}^t функции (14). Сигналы с вспомогательных выходов $L_{\tau}^{t-1}, T_{\tau}^{t-1}$, сохраняясь с предыдущего момента времени ($t - 1$), подаются, с добавлением 1 ко времени T_{τ}^{t-1} , совместно с основными сигналами на вход следующей ($\tau + 1$)-й ячейки.

Ячейка состоит из стандартных функциональных элементов:
М - схема выбора минимального напряжения с индикацией;

$R_1 \div R_4$ - схемы вычисления векторных расстояний;

$\Sigma_1 \div \Sigma_5$ - схемы сложения напряжений;

$K_1 \div K_6$ - ключевые схемы;

$\Pi_1 \div \Pi_4$ - схемы запоминания аналоговых напряжений.

Ячейка работает следующим образом. В отсутствие сигнала U_c ключи K_1, K_3 открыты, а ключи K_2, K_4 закрыты. При этом вычисляются значения L_t^t и T_t^t с одновременным их запоминанием в элементах Π_1, Π_3 . Вычисление осуществляется в соответствии с выражением (21). Три значения выражения в квадратных скобках вычисляются с помощью схем вычисления векторного расстояния R_2, R_5, R_4 и суммирующих схем $\Sigma_3, \Sigma_4, \Sigma_5$. Минимум из этих значений выбирается схемой выбора М и суммируется в Σ_2 со значением векторного расстояния $\rho^2(\xi^t, \bar{E}^t)$, вычисленным в R_1 . Результат суммирования запоминается в элементе памяти Π_1 .

Схема М, одновременно с выбором минимального из трех напряжений на входах и передачей его на выход, осуществляет индикацию входа, напряжение на котором было минимально. Индикация заключается в появлении на одном из трех дополнительных выходов I, II, III напряжения, открывающего один из ключей K_5, K_6, K_7 . В результате в элементе памяти Π_3 запоминается то значение T^m , которому соответствовал минимум выражения в квадратных скобках ф-лы (21).

Рассмотрим теперь процессы, происходящие в ячейке в моменты подачи сигнала U_c . Сигнал U_c подается в виде импульсов с частотой дискретизации параметров речевого сигнала (1). Длительность импульсов выбирается несколько большей времени установления нестационарных процессов в ячейке. При подаче сигнала U_c открываются ключи K_2, K_4 и закрываются K_1, K_3 . При этом в элементе памяти Π_2 запоминается новое значение напряжения, равное напряжению на выходе элемента памяти Π_1 , а в элементе памяти Π_4 запоминается напряжение, равное сумме напряжения на выходе Π_3 и единичного напряжения, подаваемого со входа U_o .

После исчезновения сигнала U_c ячейка переходит в состояние, являющееся исходным для следующего момента ($t + 1$) текущего времени.

На рис.8 изображена блок-схема устройства распознавания звукосочетаний. Устройство состоит из пяти различных блоков. Блок I - блок выработки p -параметрического нормализованного описания речевого сигнала (1), включающий микрофон и анализатор A_p .

Блок II - блок эталонных напряжений, соответствующих ($n + 1$) отсчетам эталона (3).

Блок III - блок однотипных ячеек, реализующих вычисление функций L_t^t и T_t^t (21) и набираемых по числу отсчетов эталона.

Блок IV - решающий блок, осуществляющий сравнение с порогом и выбор минимума функции $L(t)$ в (17).

Блок V - блок синхронизации, вырабатывающий импульсы синхронизации с частотой дискретизации параметров нормализованного описания речевого сигнала. Блок включает генератор импульсов ГИ и инвертор.

Блоки I, IV, V - общие для всех звукосочетаний. Блоки II однотипны по своей структуре и отличаются лишь величинами эталонных напряжений, выдаваемых на входы блока III. Блоки III совершенно однотипны и отличаются только количеством набранных элементарных ячеек.

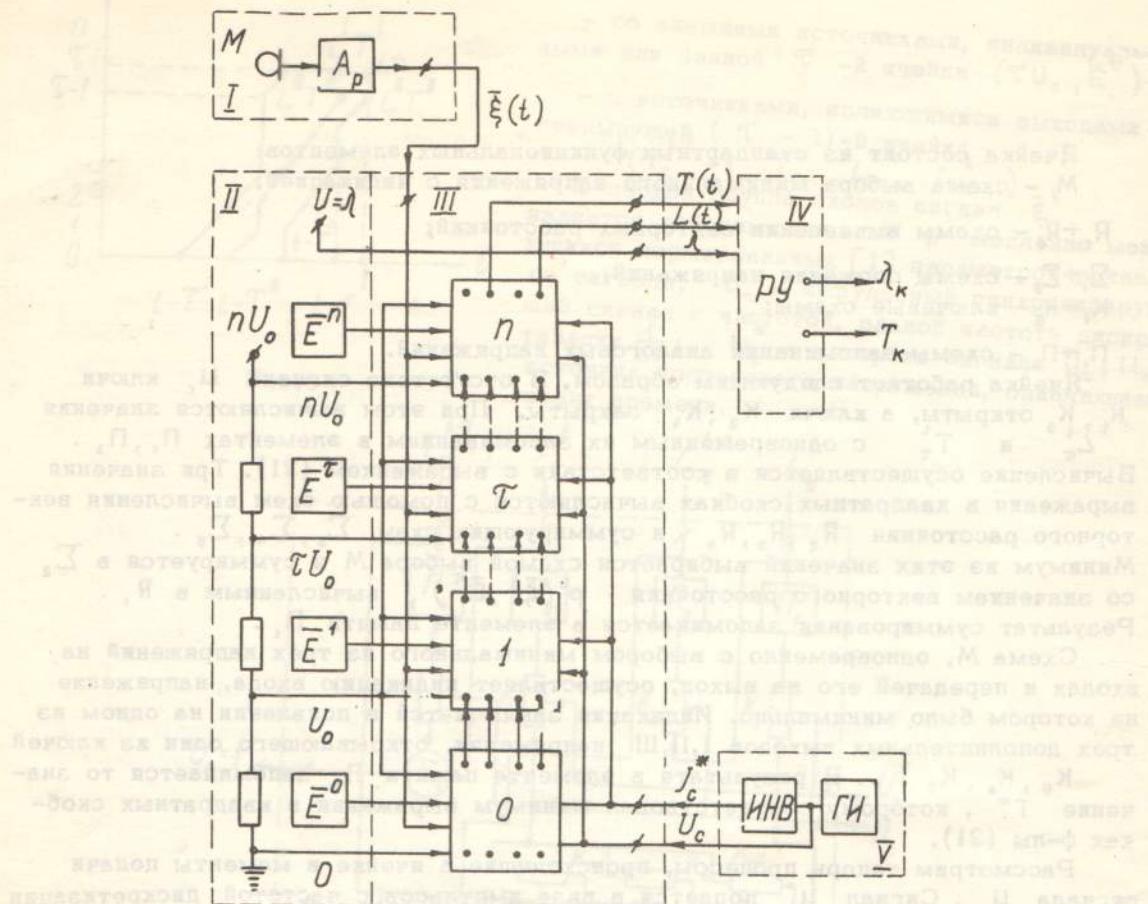


Рис.8

Таким образом, задача распознавания звукосочетаний в непрерывном потоке речи сводится к вычислению функций текущего расстояния (14) и текущей длины (15), позволяющих определить местонахождение (18) их реализаций в сигнале (1). Исполнение приведенного алгоритма принятия решения в техническом устройстве позволит вести распознавание звукосочетаний в реальном масштабе времени, что необходимо в системах синтетической телефонии и оперативного речевого управления.

ЛИТЕРАТУРА

1. А.А.Пирогов. К вопросу о фонетическом кодировании речи "Электросвязь". 1967, № 5.
2. А.А.Пирогов. Синтетическая телефония. М.,Связьиздат, 1963.
- 3.М.А.Сапожков. Речевой сигнал в кибернетике и связи. (Преобразование речи применительно к задачам техники связи и кибернетики), М.,Связьиздат, 1963.
4. А.А.Харкевич. Борьба с помехами. М.,Физматгиз, 1963,
5. В.И.Тихонов. Статистическая радиотехника. М.,"Сов.радио", 1966.
6. В.Н.Елкина, Л.С.Юдина. Статистика слов русской речи. "Вычислительные системы", Новосибирск, 1964, вып.10 и 14. (АН СССР, Сиб.отд., Ин-т математики)

7. Г.С.Служко-

Нелинейный метод анализа речевых сигналов. Труды
НИИР. 1968. № 2.

Статья поступила в редакцию 13 марта 1969 г.