

УДК 519.226:534.44

**АЛГОРИТМ КЛАССИФИКАЦИИ ТОН/ШУМ, ОСНОВАННЫЙ  
НА КРИТЕРИИ АДЕКВАТНОСТИ МОДЕЛИ АВТОРЕГРЕССИИ**

А.В.Кельманов

Целью данной работы является исследование адекватности описания моделью авторегрессии речевых сигналов при решении задачи анализа процесса речеобразования, а также построение алгоритма классификации тон/шум, основанного на критерии адекватности.

При построения моделей временных рядов обычно выделяют следующие этапы [1]: 1) постулирование общего класса моделей; 2) выбор пробной модели; 3) диагностическая проверка модели на адекватность с возвращением ко второму или первому этапу в случае неадекватности.

Предположим, что построенная в выбранном классе модель имеет минимально возможное число параметров при заданной ошибке аппроксимации. Вполне возможно, что в другом классе моделей можно получить еще более сжатое (экономичное) описание при той же ошибке аппроксимации или при том же числе параметров получить меньшую ошибку. Естественно, последняя модель предпочтительней. Кроме того, известно [1], что если ошибка аппроксимации в заданном классе моделей коррелирована, т.е. модель не полностью учитывает функциональную связь между входом и выходом, а значит, неадекватна, то существует более экономичное и точное описание. Наша цель состоит в построении не только экономичной (в заданном классе), но и адекватной модели.

При аппроксимации (в заданном классе моделей) требование адекватности, на первый взгляд, может показаться лишним. Корректное решение задачи, однако, обязательно требует проверки сделанных нами допущений. Иначе по построенной модели можно сделать ошибочные выводы. Примеры таких ошибок приводятся в [1] и будут приведены в данной работе применительно к речевому сигналу.

## §1. Об адекватности построенной модели.

В последнее время появилось много работ [4-9] по описанию речевого сигнала  $\{a_n\}$  моделью авторегрессии - проинтегрированного скользящего среднего порядка  $(p, d, q)$ :

$$\Lambda(z^{-1})w_n = B(z^{-1})\alpha_n, \quad 1 \leq n \leq N, \quad (I)$$

где  $w_n = \nabla^d a_n$ ,  $\Lambda(z^{-1}) = 1 - \sum_{i=1}^p a_i z^{-i}$  - стационарный оператор авторегрессии  $p$ -го порядка  $(AR(p))$ ;  $B(z^{-1}) = 1 - \sum_{i=1}^q b_i z^{-i}$  - оператор скользящего среднего  $q$ -го порядка;  $z^{-1}$  - оператор сдвига назад такой, что  $z^{-1}w_n = w_{n-1}$ ,  $1 = 1, 2, \dots, \max(p, q)$ ;  $\nabla = 1 - z^{-1}$  - разностный оператор со сдвигом назад;  $\alpha_n$  - нормальный белый шум с дисперсией  $\sigma_\alpha^2$ ;  $n$  - номер отсчета.

Процедуры оценивания параметров и диагностической проверки моделей достаточно хорошо разработаны [1]. Постулирование же общего класса моделей и выбор пробной модели существенно зависят от априорной информации об исследуемом временном ряде, т.е. от уровня знания физического механизма явления. Именно ошибки на первом и втором этапах приводят к тому, что модель получается неадекватной или (и) неэкономичной.

Предположим, что проводилось оценивание в классе моделей вида (I) и получена оценка параметров операторов  $\hat{\Lambda}(z^{-1})$  и  $\hat{B}(z^{-1})$ . Тогда можно показать [1], что ряд остаточных ошибок  $\hat{\alpha}_n = \hat{B}^{-1}(z^{-1}) \cdot \hat{\Lambda}(z^{-1})w_n$  для адекватной модели по мере возрастания  $N$  становится все ближе к белому шуму. Следовательно, неадекватность модели можно выявить, исследуя  $\{\hat{\alpha}_n\}$ . Иными словами, необходимо выяснять справедливость допущений относительно  $\{\alpha_n\}$ . Для этого можно воспользоваться процедурами проверки статистических гипотез.

Как правило, авторы работ [4-9], не сделав проверки модели на адекватность, переходят к ее использованию. В лучшем случае критерием адекватности является минимум среднеквадратической ошибки, т.е. критерий, который используется на этапе оценивания параметров модели и позволяет судить в основном о правильности выбора порядка модели (т.е. об экономичности в заданном классе) и при этом принципиально не позволяет сделать вывод об адекватности в смысле некоррелированности ошибки аппроксимации. В то же время, тест модели на адекватность в большинстве случаев позволяет указать на то, что именно в выбранной пробной модели неадекватно, почему неадек-

ватность возникла, как ее устранить и для каких приложений построенную модель можно использовать.

Неадекватность в смысле коррелированности ошибки аппроксимации не принципиальна лишь в том случае, если решается задача аппроксимации в заданном классе с целью сжатия описания и не оговаривается, для чего это описание будет использовано.

В речевых исследованиях традиционными задачами являются: анализ речи (выделение физических параметров, характеризующих речеобразование) [2] и распознавание.

Решение первой задачи связано с установлением соответствия между построенной математической и физической моделью (но — следущим явлением, т.е. процессом речеобразования). Поскольку взаимосвязи в физической модели описываются при помощи соответствующей математической, становится ясным, почему нужна адекватная математическая модель: если учтены не все функциональные связи в построенной модели, то и связи, существующие в физической модели, будут упущены, а, следовательно, анализ будет не точным.

Решение второй задачи связано с поиском информативной системы признаков. Такую систему обычно строят по параметрам математической модели, поэтому и в данном случае ее адекватность принципа — ливальна. В самом деле, поскольку неадекватная модель учитывает не все связи, решающее правило, построенное при помощи такой модели, также не будет их учитывать, что может привести к ошибкам распознавания. Однако известен другой подход, принятый в большинстве работ, при котором критерием адекватности является надежность распознавания. Проверка модели на адекватность (системы признаков на надежность) по результатам распознавания требует слишком больших затрат, не позволяет судить о возможности получения более высоких результатов и, наконец, вынуждает исследователя искать надежное описание методом перебора различных моделей (систем признаков), поскольку в данном случае не ясно, какие связи модель (система признаков) не учитывает. Возможно, именно этим объясняется большое число работ по выбору систем описания в задачах распознавания речи.

Рассмотрим, например, работы [6,7]. В этих работах предполагается, что ошибка аппроксимации — белый гауссовский шум и на этом основании предлагается схема максимально правдоподобного оценивания параметров моделей, несколько отличных от (I). Далее, без проверки моделей на адекватность по их параметрам рекомендуется

находить физические величины, характеризующие процесс речеобразования, а сами параметры предлагается использовать при распознавании в качестве признаков, что некорректно, поскольку предположение о нормальности и некоррелированности ошибки аппроксимация речевого сигнала предлагаемой моделью может не выполняться, или выполняться лишь частично (на отдельных участках сигнала). В последнем случае информация о неадекватности может сигнализировать о наличии различных физических механизмов генерации сигнала.

Отсоединимся на широко применяемых в настоящее время для анализа речевых сигналов моделях авторегрессии и авторегрессии в проинтегрированном виде, которые можно получить из (I), положив  $v(z^{-1}) = 1$ :

$$\Lambda(z^{-1})u_n = \alpha_n \quad (2)$$

Рассмотрим критерии адекватности применительно к данной модели.

Бартлетт [II] показал, что для любого стационарного нормального процесса  $\{u_n\}$  с нулевым средним, у которого автокорреляция

$$r_k = \frac{M[u_n u_{n+k}]}{\sigma_u^2}, \quad k = 0, 1, \dots,$$

равна нулю для задержек  $k > p$ , дисперсия выборочного коэффициента автокорреляции

$$\hat{r}_k(u) = \frac{\sum_{n=1}^{N-k} u_n u_{n+k}}{\sum_{n=1}^N u_n^2} \quad (3)$$

имеет вид:

$$\sigma^2[\hat{r}_k(u)] = \frac{1}{N} \left[ 1 + 2 \sum_{i=1}^p r_i^2(u) \right], \quad k \geq p+1 \quad (4)$$

Для белого шума  $\{\alpha_n\}$

$$r_k(\alpha) = \frac{M[\alpha_n \alpha_{n+k}]}{\sigma_\alpha^2} = 0, \quad k \geq 1,$$

поэтому из (4) получаем

$$\sigma^2[\hat{F}_k(\alpha)] \approx N^{-1}; \quad k \geq 1. \quad (5)$$

Кроме того, Андерсон [10] показал, что даже для небольших  $N$  вносочные автокорреляции  $\hat{F}_k(\alpha)$  распределены приближенно нормально с нулевым средним и дисперсией  $N^{-1}$ . Поэтому оценку статистической значимости какущихся отклонений  $\hat{F}_k(\alpha)$  от нуля при известной дисперсии можно сделать при помощи двустороннего критерия для нормально распределенных наблюдений. Необходимо проверить нулевую гипотезу  $H_0: \bar{F}(\alpha) = (r_1(\alpha), \dots, r_k(\alpha)) = 0$  ( $\forall k r_k(\alpha) = 0$ ) против альтернативной  $H_1: \bar{F}(\alpha) \neq 0$  ( $\exists k r_k(\alpha) \neq 0$ ).

Воспользовавшись таблицей интеграла вероятностей нормального распределения, для уровня значимости  $\gamma$  можно найти порог  $r_\gamma$  такой, что  $P(|\hat{F}_k(\alpha)| > r_\gamma | H_0) = \gamma$ . Поэтому нулевая гипотеза отвергается, если

$$\exists k (|\hat{F}_k(\alpha)| > r_\gamma). \quad (6)$$

Из-за того, что на практике истинные значения параметров оператора  $\Lambda(z^{-1})$  неизвестны, пользуясь (2), можно вычислить не  $\{\alpha_k\}$ , а только  $\{\hat{\alpha}_k\}$  и, как показано в [13], может оказаться рискованным придавать статистическую значимость какущимся отклонениям  $\hat{F}_k(\hat{\alpha})$  от соответствующих порогов при малых задержках  $k=1, 2, \dots, p$ . Поэтому необходимо иметь достаточно длинный ряд  $\hat{F}_k(\hat{\alpha})$ , а (6) преобразовать к виду:  $\exists k (|\hat{F}_k(\hat{\alpha})| > r_\gamma)$  для  $p < k < K$ .

На практике предпочтительнее не рассматривать отдельные значения  $\hat{F}_k(\hat{\alpha})$ , а оценить адекватность модели, например, по первым  $K$  значениям автокорреляций остаточных ошибок в совокупности. Учитывая (5) и независимость  $r_k(\alpha)$ ,  $k=1, 2, \dots, K$ , нетрудно показать, что статистика

$$N = N \sum_{k=1}^K \hat{F}_k^2(\alpha)$$

распределена примерно как  $\chi^2(K)$ . Однако эту статистику нельзя использовать для построения критерия адекватности модели  $AR(p)$  по той же причине, по какой (6) нельзя использовать в качестве критерия адекватности. Показано [13], что если модель  $AR(p)$  удовле-

творительно описывает изучаемый процесс, то статистика

$$Q = N \sum_{k=1}^K \hat{F}_k^2(\hat{\alpha}) \quad (7)$$

распределена примерно как  $\chi^2(K-p)$ . Наоборот, если модель не соответствует данным, то среднее значение  $Q$  увеличивается. Для заданного уровня значимости  $\gamma$  можно найти порог  $\chi_\gamma^2$  такой, что  $P(Q > \chi_\gamma^2 | H_0) = \gamma$ . Поэтому если  $Q > \chi_\gamma^2$ , то построенная модель неадекватна.

Еще один критерий можно построить при помощи частной автокорреляционной функции. Напомним ее определение.

Предположим, что получены оценки оператора  $AP(j)$ . Тогда при помощи метода Дурбина [1] можно получать оценки параметров для оператора  $AP(j+1)$ . Рекуррентные соотношения имеют вид:

$$\hat{a}_{j+1, i} = \hat{a}_{j, i} - \hat{a}_{j+1, j+1} \hat{a}_{j, j-1+i}, \quad i = 1, 2, \dots, j, \quad (8)$$

$$\hat{a}_{j+1, j+1} = \frac{\hat{r}_{j+1} - \sum_{i=1}^j \hat{a}_{j, i} \hat{r}_{j-1+i}}{1 - \sum_{i=1}^j \hat{a}_{j, i} \hat{r}_i}. \quad (9)$$

В этих формулах первый индекс соответствует очередному шагу при рекуррентном оценивании, второй - номеру параметра оператора для соответствующего шага, а  $\hat{r}_i$  определяется в соответствии с (3).

Величина  $a_{kk}$ , рассматриваемая как функция задержки  $k$ , называется частной автокорреляционной функцией. Она обладает тем свойством, что для процесса  $AP(p)$   $a_{kk} = 0$  при  $k > p$ .

В предположении, что порядок процесса  $AP$  равен  $p$ , в [12] показано, что выборочные частные автокорреляции для задержек  $p+1$  и выше имеют дисперсию  $\sigma^2[\hat{a}_{kk}] \approx N^{-1}$ , а в [10] показано, что  $a_{kk}$  распределены приближенно нормально с нулевым средним даже для больших  $N$ . Следовательно, решение об адекватности можно сделать при помощи частной автокорреляционной функции, используя критерий, аналогичный тому, который был построен для автокорреляционной функции остаточных ошибок.

## §2. Экспериментальные результаты

Этапы оценивания параметров и диагностической проверки моделей  $AP(p)$  при  $d = 0$  и  $I$  моделировались на ЭВМ "Минск-32". Речевой материал состоял из 10 русских слов и двух фраз: английской "Put the cap" и русской "Вода в луже медленно убывала", произнесенных одним диктором - мужчиной. Сигнал вводился в машину через семиразрядный аналого-цифровой преобразователь при различных частотах квантования ( $F_{KB}$ ): 20 кГц - для слов и 10 кГц - для фраз. Фразы исследовались в двух диапазонах частот: до 5 кГц ( $F_{KB} = 10$  кГц) и до 1 кГц ( $F_{KB} = 2$  кГц), а слова - в диапазоне частот до 1 кГц. Для уменьшения частотной полосы сигнала использовался цифровой чебышевский фильтр 3-го порядка с частотой среза 0,8 кГц и уровнем пульсации в полосе пропускания 2 дБ.

Порядок моделей выбирался таким же, каким он использовался в работах [7-9], и составил: при  $F_{KB} = 10$  кГц -  $I_2$  и  $I_4$ , а при  $F_{KB} = 2$  кГц - 4. Такие значения порядка считаются "оптимальными" [7, 9], поскольку при его дальнейшем увеличении среднеквадратичная ошибка аппроксимации почти не уменьшается.

Оценивание параметров модели проводилось асинхронно с импульсами возбуждения голосового тракта при ширине окна анализа 25,6 мсек со сдвигом от сегмента к сегменту на 16 мсек так, что при  $F_{KB} = 10$  кГц объем выборки для каждого интервала анализа составлял  $N = 256$ , а при  $F_{KB} = 2$  кГц -  $N = 52$ , и общее число проанализированных сегментов было более 500.

Значения порогов для критериев, построенных по частной автокорреляционной функции и автокорреляционной функции остаточных ошибок при уровне значимости  $\gamma = 0,0025$  для частот квантования 10 кГц и 2 кГц составляют 0,19313 и 0,42551 соответственно. Для проверки модели по критерию  $\chi^2$  использовалось 30 значений автокорреляций остаточных ошибок, поэтому при  $p = 12$  статистика (7) распределена как  $\chi^2(18)$  и при уровне значимости  $\gamma = 0,001$  имеем  $\chi^2 = 42,3$ . Для модели  $AP(4)$  имеем 26 степеней свободы и при том же уровне значимости  $\chi^2 = 54,1$ .

Визуальная (по графикам) проверка на адекватность при помощи  $\hat{F}_k(\hat{a})$  и  $\hat{a}_{kk}$  по нескольким десяткам сегментов показала, что модель  $AP(p)$  не всегда адекватно описывает речевой сигнал. На рис.1, для примера, приведены графики автокорреляционной функции сигнала  $\hat{F}_k(\hat{a})$  частной автокорреляционной функцией  $\hat{a}_{kk}$  и автокорреляционной функцией остаточных ошибок  $\hat{F}_k(\hat{a})$  для двух фонем /м/ и /х/, взятых из ан-

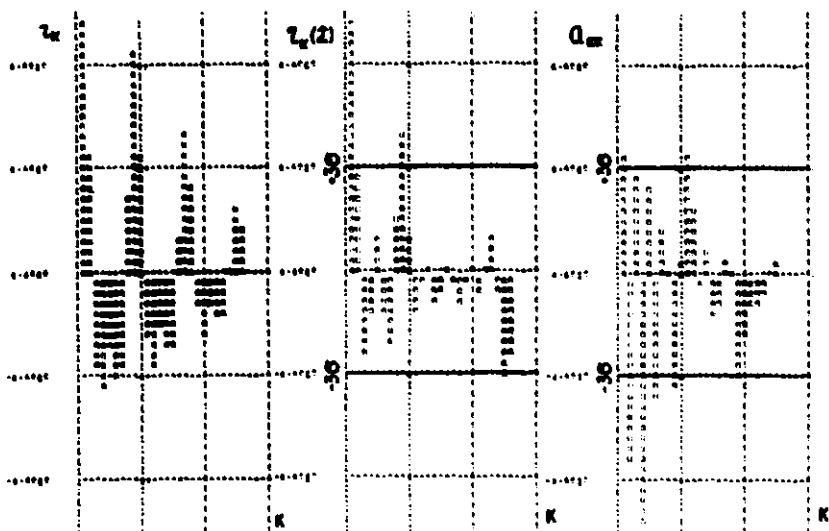


Рис. 1а. Графики функций  $r_k(s)$ ,  $r_k(\hat{\alpha})$ ,  $a_{kk}$  для модели AP (4) при частоте квантования 2 кГц для фонемы /к/, взятой из фразы "Put she can".

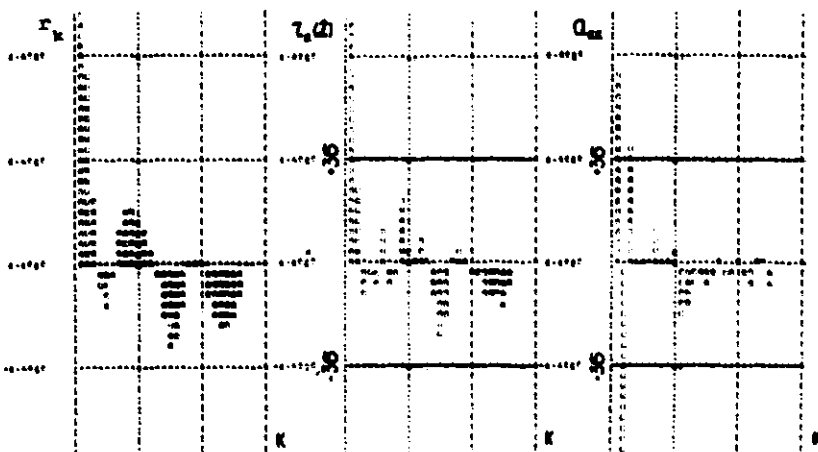


Рис. 1б. Графики функций  $r_k(s)$ ,  $r_k(\hat{\alpha})$ ,  $a_{kk}$  для модели AP (4) при частоте квантования 2 кГц для фонемы /к/, взятой из фразы "Put she can".



лейской фразы. На графиках нанесены трехсекментные интервалы. Хорошо видны отклонения от пределов значений  $\hat{F}_k(\hat{a})$  и  $\hat{S}_{k_k}$  фонемы /к / для больших задержек, в то время как для фонемы /к / графика  $\hat{F}_k(\hat{a})$  лежат в заданных пределах, а значения  $\hat{a}_{k_k}$  превышают порог лишь при малых  $k$ , где эти отклонения статистически не значимы.

Ранее отмечалось, что принимать решения по автокорреляции остаточных ошибок в частной автокорреляции не совсем удобно. Поэтому весь речевой материал был проверен по критерию  $\chi^2$ . Оказалось, что голосовые звуки описываются моделью неадекватно более чем в 98% случаев в отличие от неголосовых, для которых более чем в 99% случаев построенная модель оказалась адекватной. Это говорит о том, что данную модель нельзя использовать для выделения физических параметров речеобразования голосовых звуков. Однако критерии адекватности можно использовать для другой цели - автоматической классификации звуков речи на звонкие и глухие (тон/шум).

### §3. Алгоритм классификации тон/шум

Известные в настоящее время алгоритмы классификации тон/шум [3] могут быть разделены на два класса: алгоритмы, использующие в качестве разделяющего признака самые разнообразные физические свойства речевого сигнала, и алгоритмы, использующие вероятностные распределения, которые формулируются в виде задачи проверки статистических гипотез и, в свою очередь, могут быть разделены на два подкласса: алгоритмы, использующие непараметрические оценки плотности, и параметрические алгоритмы, в которых информация о виде распределения считается известной.

В алгоритмах параметрического типа обычно полагают, что глухие звуки представляют собой нормальный процесс с нулевым средним и известной дисперсией, в отличие от звонких, для которых вид распределения считают неизвестным. Предположение о нормальности для глухих звуков речи на практике выполняется не всегда. Кроме того, из-за недостающей информации о дисперсии, которую приходится оценивать по экспериментальным данным, решение задачи классификации, например, по критерию  $\chi^2$ , наталкивается на определенные трудности: принятие решения по теоретическому порогу приводит к существенным ошибкам, поэтому его подбирают экспериментально из компромиссных соображений [3].

Предлагаемый в настоящей работе параметрический алгоритм свободен от указанных недостатков и достаточно прост. В его осно-

не лежат рассмотренные ранее критерии диагностической проверки модели  $AP(p)$  на адекватность.

Критерий  $\chi^2$  используется в том виде, в каком он использовался при диагностической проверке. Критерий, построенный при помощи частной автокорреляционной функции, несколько изменен по той причине, что для принятия решения необходимо иметь ряд значений  $\hat{a}_{kk}$  для достаточно больших задержек и проводить много сравнений. Этого можно было бы избежать, если бы удалось найти такое  $k = k_0$ , что при заданном  $p$

$$M[\hat{a}_{k_0 k_0}] = 0 \quad (I0)$$

для неголосовых звуков и

$$M[\hat{a}_{k_0 k_0}] \neq 0 \quad (II)$$

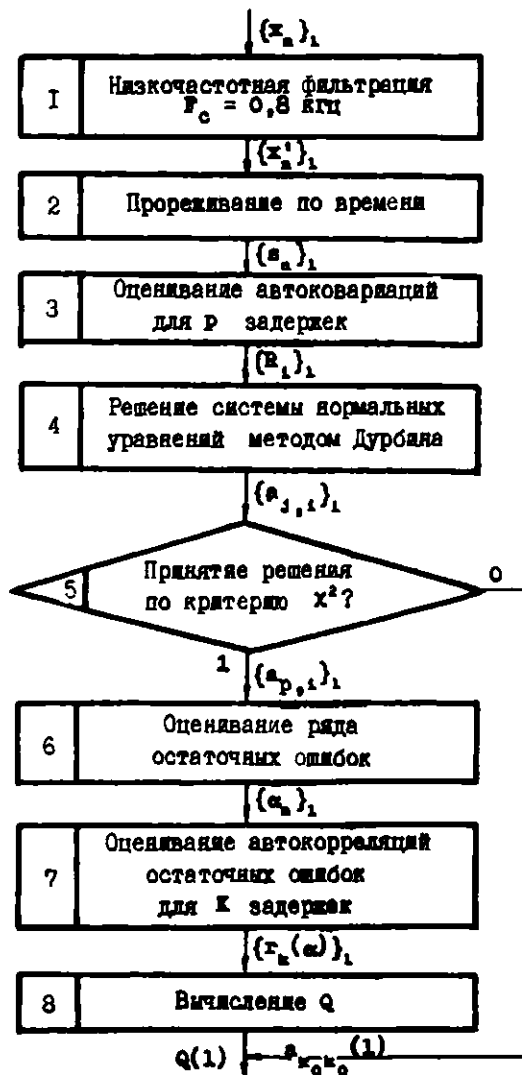
для голосовых.

Эксперименты, проведенные с речевым сигналом, ограниченным диапазоном частот до 1 кГц, показывают, что (I0) и (II) выполняются при  $p = 4$  и  $k_0 = 4$ . Поэтому для классификации можно воспользоваться значением  $\hat{a}_{44}$ .

Алгоритм классификации можно условно представить в виде блок-схемы, приведенной на рис. 2, 3. Анализируемый отрезок речевого сигнала  $\{x_n\}$ , квантованного с частотой  $F_n$ , разбивается на сегменты длительностью 25,6 мсек со сдвигом от сегмента к сегменту на 16 мсек. Для каждого сегмента 1 производится квазичастотная фильтрация при помощи чебышевского фильтра 3-го порядка с частотой среза  $F_c = 0,5$  кГц так, что диапазон частот речевого сигнала ограничивается частотой 1 кГц при частоте квантования  $F_n$ . Поэтому (при  $F_n > 2$  кГц) часть отсчетов в последовательности  $\{x'_n\}_1$  будет избыточна в соответствии с теоремой о квантовании сигнала ограниченного диапазона частот. Для перехода к частоте квантования 2 кГц из последовательности  $\{x'_n\}_1$  при помощи прореживания по времени формируется последовательность  $\{a_n\}$ , отсчеты которой берутся из  $\{x'_n\}_1$  через каждые  $F_n/2$  отсчетов. Благодаря этому сокращается число операций, необходимых для оценивания и принятия решения.

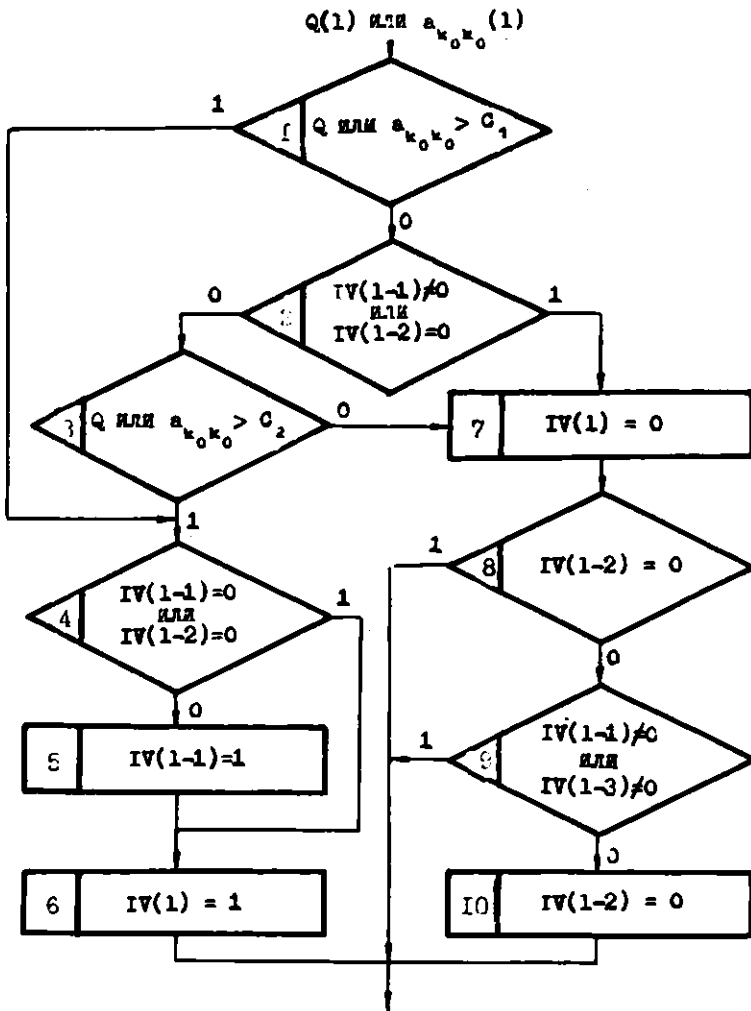
Вычисление автоковариаций ряда  $\{a_n\}$  производится по формуле

$$\hat{R}_k = \frac{1}{N} \sum_{n=1}^{N-k} a'_n a'_{n+k}, \quad k = 0, 1, \dots, p.$$



К блоку принятия решения

Рис.2. Блок-схема алгоритма. Вычисление параметров  $AR(p)$ , частной автокорреляционной функции  $a_{nk}$  и статистики  $Q$  для сегмента 1.



Вектор решений  $\langle IV \rangle = \langle 0, 0, \dots, 111, \dots, 0, 11 \rangle$

Рис. 7. Блок-схема алгоритма. Принятие решения тош/шум для сегмента 1 с коррекцией ошибочных решений для сегментов 1-1 и 1-2.

где  $a_n^i = a_n \nu_n^i$ ,  $i=1,2,\dots,N$ ,  $a \nu_n^i = 0,54 + 0,46 \cos \left[ \pi \frac{2(n-1)-N}{N} \right]$  - окно Хэмминга.

Оценивание параметров авторегрессии и частной автокорреляционной функции производится при помощи (8) и (9). Далее возможны два режима работы: принятие решения по критерию  $\chi^2$  или при помощи частной автокорреляции.

Вычисление ряда остаточных ошибок производится по формуле:

$$\hat{a}_n = \hat{\lambda}(z^{-1})a_n = a_n - \sum_{i=1}^p \hat{a}_i a_{n-i}, \quad n = 1, 2, \dots, N.$$

а оценивание их автокорреляций для  $K > p$  задержек - в соответствии с (3). Наконец, вычисление  $Q$  производим по формуле (7).

Перейдем теперь к блоку принятия решения, приведенному на рис.3. Для каждого сегмента  $l$  значения переменной  $IV(l)=0$  означает, что было принято решение "шум", если же  $IV(l)=1$  - "тон". Значения порога  $C_1$  для различных критериев при заданном уровне значимости находятся по соответствующим таблицам.

Необходимость включения в алгоритм блока принятия решения следует из того, что на переходных участках речи (тон/шум и шум/тон) возможны одиночные ошибки классификация вида: ...III... (правильное решение)  $\rightarrow$  ...IOI... (ложное) или ...OOO...  $\rightarrow$  ...OIO.... Эти ошибки можно автоматически исправить, учитывая непрерывность переходов, рассматривая решение для каждого сегмента в зависимости от решений, принятых на предыдущем и последующих сегментах.

Блок заимствован из [9] с изменениями. В частности, в упомянутой работе корректировка ...OIO...  $\rightarrow$  ...OOO..., позволяющая уменьшать число ошибок первого рода, не производится. Если наблюдаемое значение критерия меньше порога  $C_1$  и два предыдущих сегмента были голосовыми, то значение критерия сравнивается с уменьшенным порогом  $C_2 = 0,9C_1$ , т.е. для переходов тон/шум уровень значимости повышается и, таким образом, уменьшается ошибка второго рода.

Определим трудоемкость алгоритма. Оценивание автокорреляций по последовательности длины  $N$  требует  $N(p+1)$  операций сложения и умножения. Система уравнений решается за  $p^2$  арифметических операций. Вычисление остаточных ошибок производится за  $N(p+1)$  операций. Оценивание автокорреляций остаточных ошибок и вычисление  $Q$  требует  $N(K+1)K$  операций. Наконец, принятие решения осуществляется примерно за 5 операций.

Таким образом, трудоемкость алгоритма классификации по критерию  $\chi^2$  равна  $N(2r+K+3)+r^2+K+5$ , а по частной автокорреляционной функции —  $N(r+1)+r^2+5$  операций, что при  $N = 54$ ,  $r = 4$  и небольшом  $K$  позволяет принимать решение в реальном масштабе времени на ЭВМ со средней производительностью.

Полного экспериментального исследования мощности критериев не проводилось. Однако при заданном  $N$  и уровне значимости  $\gamma$ , без использования блока коррекции, ошибка второго рода для критерия  $\chi^2$  несколько ниже, чем для критерия, построенного по частной автокорреляционной функции. В какой-то мере это различие можно объяснить тем, что в первом случае для принятия решения используется  $K = 30$  значений автокорреляций остаточных ошибок, в то время как во втором — решение принимается только по одному значению частной автокорреляции.

На рис. 4, 6 приведены графики значений  $a_{i,j}, Q$  и результаты принятия решения в зависимости от номера сегмента для фразы "Pat the cat". Для сравнения алгоритмической и визуальной классификации на рис. 5 приведена картинка видимой речи, полученная по исходному сигналу, и результаты принятия решения по  $a_{i,j}$ , а на рис. 7 — картинка видимой речи, полученная по сигналу с подъемом верхних частот, и результаты принятия решения по критерию  $\chi^2$ .

Из рисунков видно, что результаты классификации для приведенных критериев несколько отличаются. Нетрудно заметить, что эти различия относятся к переходным участкам, где и решения, принятые человеком, могут быть спорными. Однако в целом классификация по критерию  $\chi^2$  более точно соответствует той, которую бы сделал обученный оператор по картинкам видимой речи.

Проведено экспериментальное сравнение алгоритма, предложенного в данной работе, с алгоритмом, описанным в [9], идея которого состоит в том, что решение тон/шум принимается по сравнению с порогом максимального значения автокорреляционной функции остаточных ошибок, вычисленной для задержек, больших чем число отсчетов в минимальном и меньших чем число отсчетов в максимальном периоде основного тона. Результаты классификации по этому алгоритму совпадают с результатами классификации при помощи частной автокорреляции. Однако трудоемкость этого алгоритма равна  $N(2r+M+3)+r^2+M+5$  операций ( $M$  — число отсчетов автокорреляционной функции остаточных ошибок), что больше трудоемкости алгоритма классификации при помощи частной автокорреляционной функции и составляет величину

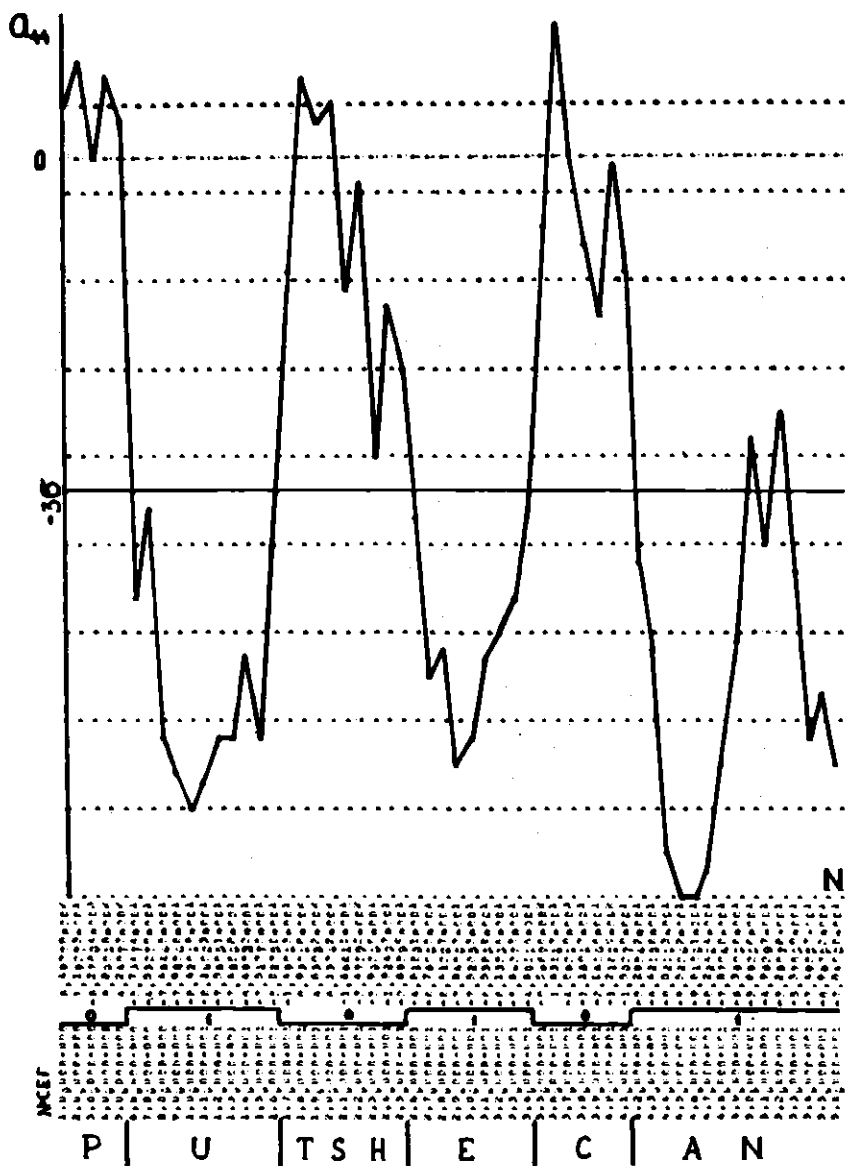


Рис. 4. График  $a_{11}(1)$  и результаты классификации по  $a_{11}(1)$  для фразы "Put she can".



Рис.5. Видимая речь для фразы "Put the cap", полученная по исходному сигналу.



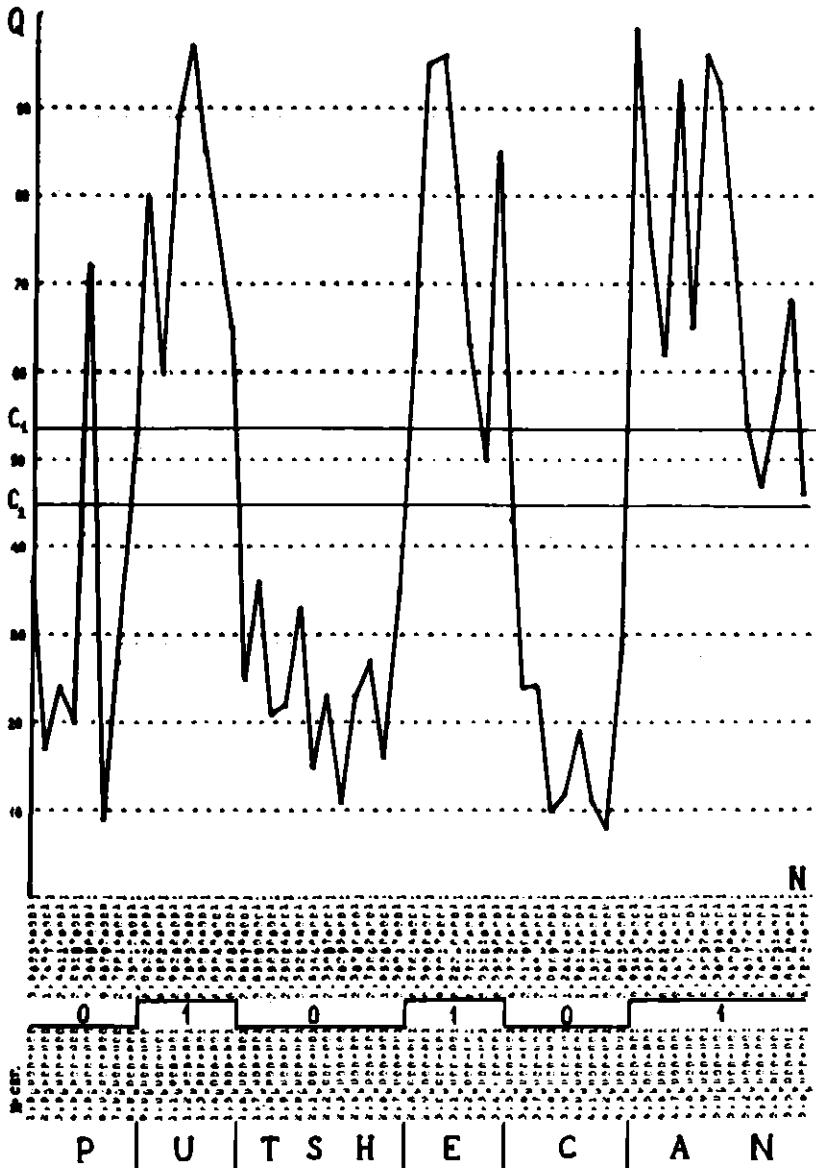


Рис. 6. График  $Q(1)$  в результате классификации по  $Q(1)$  для фразы "Put she can".

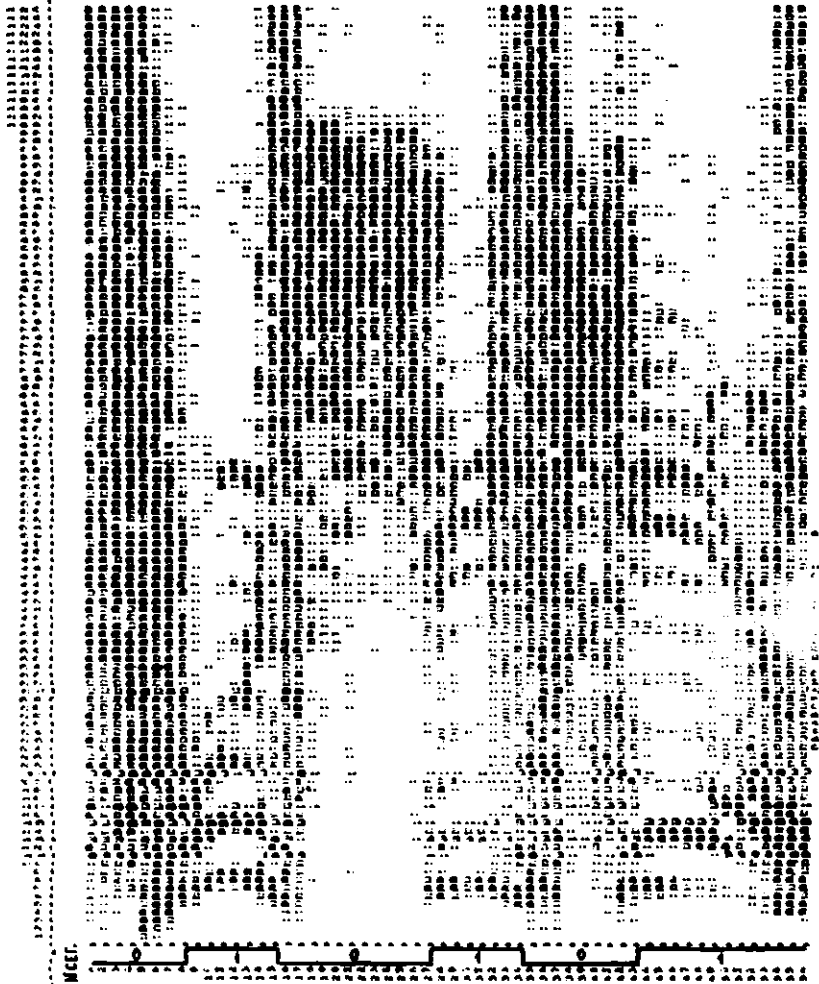


Рис. 7. Видимая речь для фразы "Put the cap", полученная по сигналу с подъемом верхних частот.

того же порядка, что и для алгоритма классификации по критерию  $x^2$ .

Описанный алгоритм был запрограммирован на языке ФОРТРАН для ЭВМ "Минск-32". Его работоспособность проверена на речевом материале, состоящем из 10 слов и одной фразы. При этом заметных расхождений с результатами визуальной классификации не наблюдалось.

Модель авторегрессии адекватно описывает неголосовые звуки более чем в 99% случаев и неадекватно голосовые — более чем в 98% случаев. Поэтому параметры данной модели нельзя использовать для выделения физических параметров речеобразования голосовых звуков речи. Тем не менее, данная модель (ее остаточные ошибки) можно использовать для других целей, например, для проведения автоматической классификации тон/шум.

Предложенный алгоритм позволяет с надежностью выше 99% разделять звонкие и глухие звуки речи в реальном масштабе времени.

#### Л и т е р а т у р а

1. БОКС Дж., ДЖЕНКИНС Г. Анализ временных рядов, прогноз и управление. М., "Мир", т. I, 1974.
2. ФАНТ Г. Акустическая теория речеобразования. М., "Наука", 1964.
3. Воккерная телефония. Под ред. А.А.Пирогова. М., "Связь", 1974.
4. ЛЮЗОВСКИЙ В.С. Аппроксимация отклика системы в  $x$ -плоскости и формантный анализ речи. — В кн.: Вычислительные системы, вып. 37, Новосибирск, 1969, с. 22—37.
5. ВИННИК Г.К., ЛЮДОВИК Е.К. Оценка параметров речевого тракта как устойчивой линейной стохастической системы. — В кн.: Распознавание образов, Киев, 1975, с. 3—29.
6. ЛЮДОВИК Е.К. Совместное оценивание нуль-полосных параметров речевого тракта и характеристик источников возбуждения. — В кн.: Обработка и распознавание сигналов. Киев, 1975, с. 121—134.
7. МАКУЛ Дж. Линейное предсказание. Обзор. — "Труды института инженеров по электротехнике и радиоэлектронике", 1975, т. 63, № 4, с. 20—44.
8. ATAL B.S., HANAUER S.L. Speech analysis and synthesis by linear prediction of the speech wave. — "J. Acoust. Soc. Amer.", 1971, v. 50, № 2, pt 2, p. 637—655.
9. MARKEL J.D. Application of digital inverse filter for automatic formant and  $F_0$  analysis. — "IEEE Trans. Audio Electroacoust.", 1973, v. AU-21, № 3, p. 154—160.

10. ANDERSON R.L. Distribution of the serial correlation coefficient.- "Ann.Math.Statistics", 1942, v.13, N 1, p.1-13.

11. BARTLETT M.S. On the theoretical specification properties of sampling properties of autocorrelated time series.- "J.Roy. statist.Soc.", 1946, Series B, v.8, N 1, p.27-41.

12. QUENOUILLE M.H. Approximate test of correlation in time series.- "J.Roy.Statist.Soc.", 1949, Series B, v.11, N 2, p.68-83.

13. BOX G.E., PIERCE D.A. Distribution of residual autocorrelations in autoregressive - integrated moving average time series models.- "J.Amer.Statist.Assoc.", 1970, v.65, N 332, p.1509-1526.

Поступила в ред.-изд.отд.  
29 марта 1978 года