УДК 621.36.2

РАСПОЗНАВАНИЕ РЕЧЕВОГО СИГНАЛА НА ФОНЕ БЕЛОГО ШУМА И УЗКОПОЛОСНОЙ ПОМЕХИ

И.Н. ПРЕСНЯКОВ, С.В. ОМЕЛЬЧЕНКО

Предлагается новый подход к решению задачи адаптивного распознавания речи на основе модифицированного метода модели авторегрессии скользящего среднего (АРСС). Проведенные экспериментальные исследования на реальных сигналах доказывают возможность приемлемого качества распознавания речи этим методом.

A new approach for solving a problem of adaptive identification of a speech on the basis of a modified method of autoregression moving average model. The carried out experimental investigation for real signals prove the possibility of reasonable quality of the speech identification by this method.

ВВЕДЕНИЕ

На точность оценки параметров моделей линейного предсказания существенно влияет уровень помех, возникающих при передаче речи по телефону. Другой источник помех — это акустические шумы, связанные с работой приборов. Задача этих исследований состоит в определении метода снижения влияния шумов на точность систем автоматического распознавания слов речи. Высокая чувствительность оценок параметров авторегрессии к широкополосному белому шуму вынуждает разработчиков систем распознавания и обработки речи проводить исследования, направленные на создание методов, снижающих влияние шумов на точность линейного предсказания.

Целью данной статьи является построение алгоритмов распознавания речи в условиях действия помех.

1. ПОСТАНОВКА ЗАДАЧИ

Полагается, что на вход системы распознавания поступает временная последовательность отсчетов речевого сигнала x_n, $n=\overline{1,N}$, взятых с интервалом дискретизации Δt.

Для создания алгоритмов распознавания важны априорные сведения о вводимых словах.

Эталоны слов для каждого из дикторов заданы в виде классифицированных обучающих выборок.

Считается, что время предъявления речевых единиц в речевом сигнале априори неизвестно. Положим, что априорные вероятности предъявления для всех структурных речевых единиц одного типа одинаковы.

Необходимо построить алгоритм, который по предъявленной реализации речи выносит решение о конкретном слове и обеспечивает максимум средней вероятности правильного распознавания слов, а также удовлетворяет ограничениям на среднюю вероятность правильного распознавания слов речи при воздействии аддитивной помехи в канале связи с заданным отношением сигнал-шум q.

Голосовой аппарат человека представляет собой акустическую систему, состоящую из ротового и носового каналов, возбуждаемую квазипериодическими импульсными колебаниями голосовых связок и турбулентным шумом. Турбулентный шум образуется путем проталкивания воздуха через сужения в определенных областях голосового тракта. Голосовой аппарат, возбуждаемый указанными источниками, действует как линейный фильтр с изменяющимися во времени параметрами, на выходе которого формируется речевой сигнал.

В передаточной функции речеобразующей системы для носовых звуков, например «м», «н», наряду с резонансами появляются антирезонансы (нули) [5]. Обычно при распознавании речи используют модель авторегрессии (АР), которая адекватно описывает положение полюсов. Адекватной моделью, хорошо описывающей положение нулей, является модель скользящего среднего (СС).

В качестве модели цифрового фильтра голосового тракта выбрана модель АРСС, которая хорошо описывает положение полюсов и нулей.

Полагается, что речевой сигнал в канале связи аддитивно взаимодействует с помехой, которая также описывается моделью АРСС.

Различие соотношений сигнал-шум при формировании эталонов слов речи и на этапе распознавания приводит к ухудшению качества распознавания слов речи, либо невозможности распознавания слов речи.

В статье показано, что этот недостаток может быть устранен на основе адаптивного метода, включающего оценивание отношения сигнал-шум.

Решение задачи распознавания речи и ряда других задач во многом связано с успешным проведением сегментации речи на речевые единицы. Подобная сегментация на этапе распознавания речевых единиц позво-

ляет исключить избыточные процедуры принятия решений по сигналам, не несущим речевую информацию. Задачи сегментации состоят в членении речи на структурные единицы и оценивании их временных границ. Некоторые алгоритмы сегментации описаны и исследованы в [1–4].

2. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА РЕЧЕВОГО СИГНАЛА

Рассмотрим предварительную обработку речевого сигнала цифровым фильтром, построенным на основе модели АРСС. Такой фильтр необходим для исключения коррелированных помех из сигнала и выравнивания АЧХ распознаваемых сигналов. Эффективность выбеливания для повышения качества распознавания и сегментации слов речи показана в работах [2–4]. Полагается, что априори известен интервал времени, в течение которого отсутствует речь (пауза). Такой интервал времени используется для оценивания АРСС-параметров фильтра предварительной обработки.

Известно, что для оценивания АРСС-параметров применяются процедуры раздельного оценивания АР- и СС-параметров [7]. Сначала оцениваются АР-параметры, а затем их оценки используют для построения обратного фильтра, который будет применен к исходным данным. Последовательность остаточных ошибок на выходе этого фильтра должна характеризовать процесс скользящего среднего, к которому будет применена процедура оценивания СС-параметров.

Наши исследования на примере с узкополосной помехой показали, что раздельное оценивание АР- параметров в условиях действия белого шума приводит к ухудшению качества спектральных оценок параметров выбеливающего фильтра (смещается и расширятся полоса фильтра). Экспериментально показано, что точность АР-параметров можно повысить за счет коррекции корреляционной функции с учетом уровня белого шума.

Модель АР описывается разностным уравнением

$$n_t = \sum_{u=1}^{p} a_u n_{t-u} + \xi_t , \qquad (1)$$

где a_u — коэффициенты АР; p — порядок модели АР; ξ_t — некоррелированные ошибки предсказания.

Минимизируя дискретную ошибку предсказания по параметру a_u, приходим к уравнению Юла — Уокера:

$$[r]\cdot\bar{a}=\bar{r},$$

где матрицы и векторы, входящие в уравнение, имеют вид:

$$[r]=\begin{bmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ \dots & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{bmatrix}, \quad \bar{r}=\begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_p \end{bmatrix}, \quad \bar{a}=\begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix}.$$

Корреляционная матрица представлена компонентами

$$r_j = R_j / R_0 .$$

$$R_{nj} = \frac{1}{(T+1-j)\cdot(L2+1-L1)} \sum_{v=L1}^{L2} \sum_{i=0}^{T-j} (s_{i+j}^{(v)} \cdot s_i^{(v)})$$ —

оценка корреляционной функции сигнала в паузе, v — номер выборки.

Процедура оценивания дисперсии аддитивного белого шума затруднена наличием узкополосной помехи.

При наличии аддитивного белого шума и узкополосной помехи (будем считать их статистически независимыми) сигнал в паузе описывается выражением

$$y_t = x_{yt} + n_{\delta t}$$

с корреляционной функцией

$$R_{yj} = R_{nyj} + D_n \cdot \delta(j) ,$$

где R_j^x — корреляционная функция сигнала при отсутствии шума; D_n — дисперсия белого шума; $\delta(j)$ — дельта-функция Дирака.

Поэтому корреляционная функция узкополосной помехи в паузе корректируется с учетом уровня белого шума

$$R_{yj} = R_{nyj} + D_n \, \delta(j) ,$$

где $\delta(j)=\begin{cases}1, & \text{если } j=0\\ 0, & \text{если } j\neq 0\end{cases}$.

Приближенные оценки дисперсии белого шума D вычисляются по спектральным оценкам шума в паузе $S^{(v)}(i)$ в виде

$$D = \min(D_1, D_2, \dots, D_J) ,$$

где $D_j = \frac{1}{\Delta\cdot(L2+1-L1)} \sum_{v=L1}^{L2} \sum_{i=(j-1)\Delta+1}^{j\Delta} (S^{(v)}(i) \cdot S^{(v)}(i))$ — оценки дисперсии шума, построенные в i-й полосе частот.

Вектор оценок коэффициентов АР находится из выражения

$$\bar{a} = [r]^{-1} \cdot \bar{r} .$$

Алгоритм оценивания ошибки предсказания описывается выражением

$$y_t = x_t - \sum_{u=1}^{p} \hat{a}_u x_{t-u} , \qquad (2)$$

где $\hat{a}_u$ — оценки коэффициентов АР.

Оценка нормированной корреляционной функции ошибки предсказания сигнала в паузе

$$K_{yj} = \frac{1}{(T+1-j)\cdot(L2+1-L1)} \sum_{v=L1}^{L2} \sum_{i=0}^{T-j} (y_{i+j}^{(v)} \cdot y_i^{(v)}) , \qquad (3)$$

где v — номер выборки, T — период наблюдения.

Фильтрация сигнала ошибки предсказания описывается разностным уравнением

$$s_t = -\sum_{u=1}^{p} b_u s_{t-u} + y_t , \qquad (4)$$

где $b_u = K_{yu} / K_{y0}$ — коэффициенты фильтра, являющиеся результатом оценивания нормированной корреляционной функции ошибки предсказания.

Нормированная АЧХ фильтра

$$H(n2\pi/T)=\frac{|1-\sum_{k=1}^{p}(a_k\cdot e^{-ikn2\pi/T})|}{|\sum_{k=0}^{p}(b_k\cdot e^{-ikn2\pi/T})|}.$$

Коэффициенты $\vec{a}=(a_0,a_1,....,a_p)$ и $\vec{b}=(b_0,b_1,....,b_p)$ выбеливающего АРСС фильтра вычисляют с использованием выборок речевого сигнала, взятых в период молчания.

Анализ результатов статистических исследований показывает, что предлагаемый метод уточнения параметров модели значительно улучшает спектральные оценки случайных процессов в шумах.

На рис. 1 показаны спектры реализаций, полученные для случаев: *а* — смеси узкополосной случайной помехи с центральной частотой f_0=1100 Гц, полосой Δf =100 Гц и белого шума; *б* — результата подавления помехи АР фильтром; *в* — результата подавления помехи АРСС фильтром с коррекцией корреляционной функции, используемой для оценивания АР-параметров.

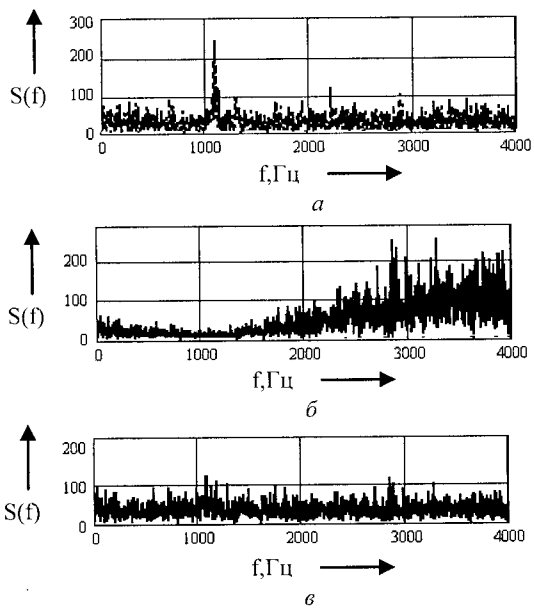


Рис. 1. Спектры реализаций: *а* — смеси узкополосной случайной помехи и белого шума; *б* — результата подавления помехи АР фильтром; *в* — результата подавления помехи АРСС фильтром с коррекцией корреляционной функции

На рис. 2, *а* и *б* приведены корреляционные функции шума после выбеливания помехи АР фильтром и фильтром АРСС соответственно. Из рисунка видно, что фильтр АРСС выполняет декорреляцию лучше, чем фильтр АР.

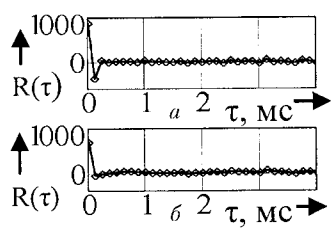


Рис. 2. Корреляционная функция шума в паузе после фильтрации: *а* — АР; *б* — АРСС

На рис. 3, *а* и *б* приведены оценки амплитудно-частотных характеристик выбеливающего фильтра с передаточной характеристикой $H(m)$ для АР фильтра и фильтра АРСС соответственно (узкополосная помеха с центральной частотой f_0=2500 Гц и полосой Δf =100 Гц).

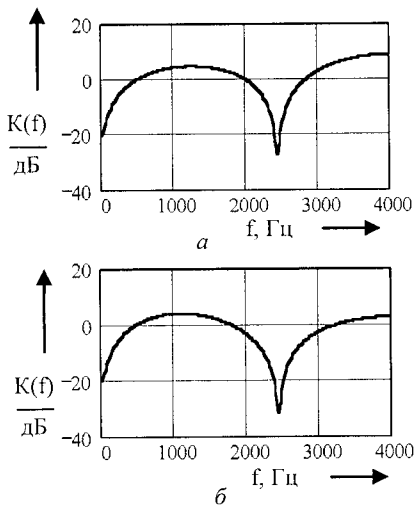


Рис. 3. АЧХ фильтра предварительной обработки речи: *а* — АР; *б* — АРСС

На рис. 4, *а* приведена сонограмма слова «четыре» при действии узкополосной помехи с соотношением сигнал-шум по мощности $q^2=1$ и частотой 2500 Гц, а на рис. 4, *б* — после обработки АРСС фильтром. Проведенные исследования подтверждают эффективность алгоритма предварительной обработки (2, 4) в условиях действия узкополосных случайных помех (рис. 4).

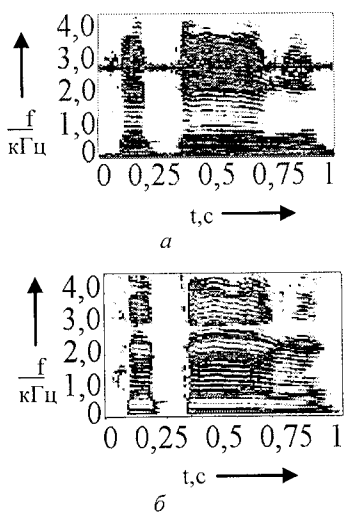


Рис. 4. Сонограмма слова «четыре»: *а* — с помехой , *б* — после обработки

3. ВЫБОР АЛГОРИТМА РАСПОЗНАВАНИЯ

После предварительной обработки выполняется вычисление признаков, необходимых для распознания слов речи.

По каждой v-й выборке оценивается корреляционная функция

$$R^{(v)}{}_{yj} = \frac{1}{(L2+1-L1)} \sum_{i=0}^{T-j} (y_{i+j}^{(v)} \cdot y_i^{(v)}).$$

Корреляционная функция полученных эталонов может пересчитываться с учетом статистических свойств шума на этапе распознавания.

При действии белого шума высокого уровня на входе устройства скорректированная корреляционная матрица эталона

$$R = R_э + I(f(\sigma_c{}^2) - \sigma_э{}^2),$$

где $R_э$ — исходная корреляционная матрица; $f(\sigma_c{}^2) \approx \sigma_c{}^2/(\hat{q}/10)$ — корректирующая функция; I — единичная матрица; $\sigma_c{}^2$, $\sigma_э{}^2$ — дисперсия сигнала на входе устройства распознавания и эталона соответственно.

Методом Левинсона или Дарбина вычисляются оценки коэффициентов авторегрессии.

Оценивание ошибки предсказания выполняется в соответствии с алгоритмом

$$y_{t,u,v}{}^{(k,j)} = x_{t,u}{}^{(k)} - \sum_{\tau=1}^{p} \hat{a}_{e\tau,v}{}^{(j)} x_{t-\tau,u}{}^{(k)},$$

где $x_{t,u}{}^{(k)}$ — результат выбеливания фильтром предварительной обработки на этапе обучения.

Далее производится оценивание корреляционной функции ошибки предсказания и её коррекция с учетом уровня помех

Таким образом, производится оценивание параметров

$$\vec{a}_{em}{}^{(j)} = (a_{em0}{}^{(j)}, a_{em1}{}^{(j)}, \ldots, a_{emp}{}^{(j)}) \text{ и}$$

$$\vec{b}_{em}{}^{(j)} = (b_{em0}{}^{(j)}, b_{em1}{}^{(j)}, \ldots, b_{emp}{}^{(j)})$$

обеляющего фильтра для каждого v-го блока речевого сигнала с учетом помеховой обстановки на этапе распознавания.

Для каждого блока производится нормирование распознаваемого речевого сигнала

$$x_{t,u}{}^{(k)} = x_{t,u}{}^{(k)} / (\sum_{\tau=1}^{T} x_{\tau,u}{}^{(k)2} / T)^{1/2},$$

где $x_{t,u}{}^{(k)}$ — результат выбеливания фильтром предварительной обработки на этапе распознавания.

Алгоритм оценивания результата АРСС фильтрации на этапе распознавания описывается двумя разностными уравнениями:

$$y_{t,u,v}{}^{(k,j)} = x_{t,u}{}^{(k)} - \sum_{\tau=1}^{p} \hat{a}_{e\tau,v}{}^{(j)} x_{t-\tau,u}{}^{(k)},$$

$$n_{t,u,v}{}^{(k,j)} = -\sum_{\tau=1}^{p} b_{e\tau,v}{}^{(j)} n_{t-\tau,u,v}{}^{(k,j)} + y_{t,u,v}{}^{(k,j)}. \quad (5)$$

В случае фильтрации на основе АР, коэффициенты $b_{e\tau,v}{}^{(j)} = 0$, поэтому второе уравнение (5) имеет вид $n_{t,u,v}{}^{(k,j)} = y_{t,u,v}{}^{(k,j)}$.

Решение для k-го сегмента о наличии заданной речевой единицы выносится в соответствии с выражением

$$i(k) = \arg \max_{j \in [1,M]} R^{(k,j)}, \quad (6)$$

здесь усредненная мера

$$R^{(k,j)} = \sum_{m=1}^{M} \sum_{v=1}^{V} \sum_{h=H_1}^{H_2} 1/D(n_{t,v+h,v}{}^{(k,j,m)}),$$

где

$$D(n_{t,v+h,v}{}^{(k,j,m)}) = \sum_{t=1}^{N} n_{t,v+h,v}{}^{(k,j,m)2}/N -$$

$$-(\sum_{t=1}^{N} n_{t,v+h,v}{}^{(k,j,m)}/N)^2$$

— оценка дисперсии сигнала с выхода АРСС фильтра для одного блока; M — количество эталонов.

Рассмотрим решение задачи распознавания речи с использованием процедур динамического программирования (ДП). Процедура динамического программирования служит для нелинейной временной нормализации.

Решение о слове принимается по максимуму меры сходства распознаваемого сегмента слова к одному из слов эталона. В качестве меры расхождения между векторами признаков вводится расстояние между ними, например, в виде

$$d^{(k,n)}(m,v) = 1/\sum_{t=1}^{N} n_{t,v}{}^{(k,n,m)2}.$$

Минимальное значение $D(A)$ достигается при оптимальном согласовании временных расхождений между образами.

Вычисление начинается от концов сегментов слов и завершается к их началу. Рекуррентное уравнение алгоритма представляется следующим образом:

$$D^{(k,n)}(i,j) = \max \begin{cases} D^{(k,n)}(i-1,j-2) + 2d^{(k,n)}(i,j-1) + d^{(k,n)}(i,j); \\ D^{(k,n)}(i-1,j-1) + 2d^{(k,n)}(i,j); \\ D^{(k,n)}(i-2,j-1) + 2d^{(k,n)}(i-1,j) + d^{(k,n)}(i,j). \end{cases} \quad (7)$$

Альтернативный рекуррентный алгоритм вычисления уравнения представляется следующим образом:

$$D^{(k,n)}(i,j) = \max \begin{cases} D^{(k,n)}(i,j-1) + \chi \cdot d^{(k,n)}(i,j); \\ D^{(k,n)}(i-1,j-1) + d^{(k,n)}(i,j); \\ D^{(k,n)}(i-1,j) + \chi \cdot d^{(k,n)}(i,j). \end{cases} \quad (8)$$

Решение для k-го сегмента о наличии заданной речевой единицы принимается по правилу

$$i(k) = \arg \max_{j \in [1,M]} D^{(k,n)}(0,0). \quad (9)$$

Рассмотрим другие алгоритмы распознавания, использующие АРСС фильтры. Принятие решение для

k-го сегмента о слове выносится в соответствии с выражением

$$i(k) = \arg\max_{j\in[1,M]} p(k \mid j),$$

здесь условная вероятность

$$p(k \mid j) = \sum_{m=1}^{M}\prod_{v=1}^{V}\sum_{h=H_1}^{H_2}\frac{1}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{\sum_{t=1}^{N} n_{t,v+h,v}^{(k,j,m)2}}{2\sigma_v^{\,2}}\right),$$

где M — количество эталонов.

Оценки формантных частот спектрально-полосным методом могут вычисляться как среднеэффективные частоты в заданных полосах частот с соответствующего выхода полосового фильтра с заданными граничными частотами $f_в^{(m)}$ и $f_н^{(m)}$, указанными в табл. 1 для каждого из блоков.

Таблица 1

m	$f_н^{(m)}$, Гц	$f_в^{(m)}$, Гц
1	200	850
2	850	2200
3	2200	3000
4	3000	4000

Адаптивные оценки формантных частот с учетом разной помеховой обстановки на этапе обучения и распознавания строятся следующим образом. Перед распознаванием и обучением проводится нормировка всего сигнала по мощности помехи в паузе. На этапе обучения и распознавания вычисляются формантные частоты как среднеэффективные частоты в заданных полосах частот

$$\hat{f}_э^{(m)} = \sum_{i=f_н^{(m)}}^{f_в^{(m)}} i\left(S_{э,i}^2 + A(\hat{q}_{тек})\right) / \sum_{i=f_н^{(m)}}^{f_в^{(m)}} \left(S_{э,i}^2 + A(\hat{q}_{тек})\right),$$

а на этапе распознавания

$$\hat{f}_{тек}^{\ (m)} = \frac{\sum_{i=f_н^{(m)}}^{f_в^{(m)}} i\left(S_i^2 + G_i C\left(1-\mathrm{sgn}(q_{тек}-1)\right)\right)}{\sum_{i=f_н^{(m)}}^{f_в^{(m)}} \left(S_i^2 + G_i C\left(1-\mathrm{sgn}(q_{тек}-1)\right)\right)},$$

где ($f_в^{(m)}$, $f_н^{(m)}$) — диапазон частот для m-й форманты;

S_2 — вычисленный дискретный спектр;

$$\hat{q} = \frac{\sum_{i=\tau_{c,1}}^{\tau_{c,2}} (S_i)^2/(\tau_{c,2}-\tau_{c,1}+1)}{\sum_{i=\tau_{n,1}}^{\tau_{n,2}} (S_i)^2/(\tau_{n,2}-\tau_{n,1}+1)} -$$

оценка отношения сигнал-шум на этапе обучения и распознавания соответственно;

$$A(\hat{q}_{тек}) = G_i(B/\hat{q}_{тек}^{\ 2} + C(1-\mathrm{sgn}(\hat{q}_{тек}-1)) -$$

корректирующая функция; G_i — весовая функция с учетом спектра помехи и характеристики фильтра предварительной обработки речевого сигнала; функция

$$\mathrm{sgn}(u) = \begin{cases} 1, & \text{если } u \ge 0, \\ 0, & \text{если } u < 0. \end{cases}$$

Улучшить точность первичного оценивания траектории формант можно путем выполнения операции сглаживания $\hat{f}_{ср}^{\ (m)} = \sum_{r=-v}^{u} \hat{f}^{(m-r)} W_r$, где $\sum_{r=-v}^{u} W_r = 1$.

Процедура вычисления формант может быть повторена, но при этом в качестве граничных полос частот используют

$$\hat{f}_в^{(m)} = \hat{f}^{(m)} + \Delta,\ \hat{f}_н^{(m)} = \hat{f}^{(m)} - \Delta, \qquad (10)$$

где $\hat{f}^{(m)}$ — форманты, вычисленные на предыдущем этапе; Δ — границы диапазона поиска формант. Простейшей среди рекуррентных процедур является двухэтапная.

Статистики $R_{\text{пр}}^{(l,s,d)}$ вычисляют по оценкам формантных циклических частот $\hat{\omega}_{i,\text{пр}}(n)$ и $\hat{\omega}_{i+j,\text{пр}}^{l,s}(n)$, а $R_{\text{обр}}^{(l,s,d)}$ по оценкам $\hat{\omega}_{i,\text{обр}}(n)$ и $\hat{\omega}_{i+j,\text{обр}}^{l,s}(n)$ в виде

$$R^{(l,s,d)} = \sum_{d=1}^{D}\sum_{s=0}^{S(d)} P_в(\omega_{s,d}) \prod_{n=1}^{N}\prod_{i=1}^{L(n)}\sum_{j=-J}^{J}\sum_{h=-H}^{H} P_c(\omega_{j,h}) \times$$

$$\times \frac{c_i(n)}{\pi(c_i(n)^2 + |\hat{\omega}_i(n) - \hat{\omega}_{i+j}^{l,s}(n+h)|^2)},$$

где $c_i(n) = \Delta\omega_i(n)/2$ — параметр, $\Delta\omega$ — циклическая полоса частот формант.

Ширина полос формант изменяется в зависимости от произносимых фонем в пределах от 50 до 400 Гц. В некоторых случаях можно положить, что полосы формант одинаковы, тогда $c_{i,пр}(n) = \Delta\omega/2$.

Решение о слове на основе модели АРСС и спектрально-полосных оценок формантных частот принимается из условия

$$i = \arg\max_{l=\overline{0,M}} \left(\sum_{d=1}^{D}\sum_{s=0}^{S(d)} (k_{SP}\cdot \ln(R_{SP}^{(l,s,d)}) + k\cdot R^{(l,s,d)}\right),$$

где $R_{SP}^{(l,s,d)}$ — функционал, построенный на основе спектрально-полосных оценок.

4. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Испытания приведенных выше алгоритмов распознавания слов проводились на основе данных, введенных в ЭВМ с микрофона через звуковой интерфейс с частотой дискретизации $F_д$=8 кГц.

Качество распознавания сигналов оценивалось средней вероятностью правильного распознавания, которая получалась на контрольных выборках реализаций методом статистических испытаний.

Для проверки предложенного алгоритма (6) распознавания речи на основе оценки параметров процесса АР были проведены экспериментальные исследования алгоритмов распознавания речи. С помощью генератора АР получены выборки процесса АР с задаваемыми центральными частотами и ширинами полос. Применялось усреднение по 20 выборкам из генеральной совокупности длиной 256 отсчетов каждая.

Алгоритм (6) позволил получать при действии узкополосной помехи с центральной частотой f, полосой Δf =100 Гц и соотношением сигнал-шум по мощности $q^2 = 1$ среднюю вероятность правильного распознавания десяти слов речи (десять цифр) от 0,5 до 0,7 (рис. 5), а после обработки АРСС фильтром — P=0,91. Для соотношения сигнал-шум по мощности $q^2 = 0,1$ с центральной частотой 1500 Гц и полосой Δf =100 Гц получена средняя вероятность правильного распознавания десяти слов речи после обработки АРСС фильтром 0,3, а после обработки АРСС фильтром и адаптивным алгоритмом распознавания — 0,8. Распознавание слов речи при тех же параметрах помех невозможно для формантных алгоритмов, описанных в [1].

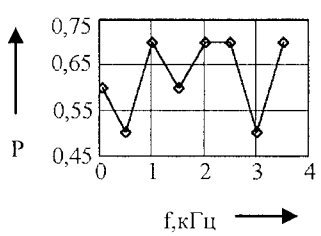


Рис. 5. Зависимость вероятности правильного распознавания P (без выбеливания, $q^2 = 0$ дБ) от центральной частоты помехи f

На рис. 6. приведена зависимость вероятности правильного распознавания P от отношения сигнал-шум q для эталона с отношением сигнал-шум: $1-q^2 = 11$ дБ; $2-q^2 = 5$ дБ.

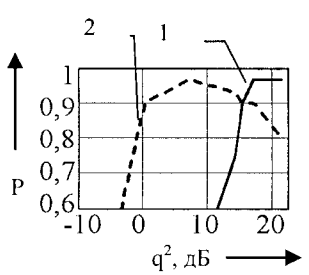


Рис. 6. Зависимость P от q^2 для эталона с: $1 - q^2 = 11$ дБ; $2 - q^2 = 5$ дБ.

На рис. 7 приведена зависимость вероятности правильного распознавания P от отношения сигнал-шум q алгоритмов с перерасчетом корреляционной функции к отношению сигнал-шум: $1 - q^2 = 3,5$ дБ $2 - q^2 = -6$ дБ; $3 - q^2 = -26$ дБ.

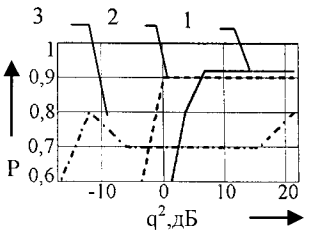


Рис. 7. Зависимость вероятности правильного распознавания P от соотношений сигнал-шум q^2 алгоритмов с коррекцией корреляционной функции при: $1 - q^2 = 3,5$ дБ; $2-q^2 = -6$ дБ; $3-q^2 = -26$ дБ.

Как показали экспериментальные исследования (рис. 8), коррекция корреляционных функций эталонов при $0,1 \le \hat{q} \le 10$ должна быть выполнена с вычислением расчетных отношений сигнал-шум q_p по формуле

$$q_p \approx \hat{q} \cdot \sqrt{\hat{q}/10}\,.$$

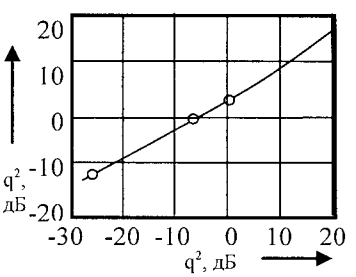


Рис. 8. Зависимость расчетного отношения сигнал-шум q^2_p от оценки отношения сигнал-шум q^2

Предложенные алгоритмы распознавания слов речи на основе модели АРСС позволили получать при отсутствии помех и одном эталоне на каждое из десяти слов вероятность правильного распознавания 0,92 а при трех эталонах на каждое слово — вероятность правильного распознавания 0,96.

Установлено, что для алгоритмов на основе модели АРСС и полосы частот речевого сигнала 4 кГц наибольшая вероятность правильного распознавания слов будет для 11–13-го порядка (рис. 9).

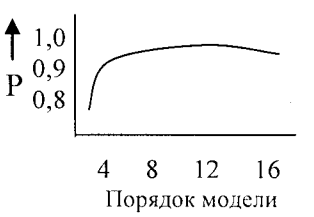


Рис. 9. Зависимость вероятности правильного распознавания слов от порядка модели

Применение комбинации спектрально-полосных методов и фильтров АРСС позволяет осуществлять распознавание слов речи с вероятностью правильного распознавания 0,97, что выше, чем при формантных методах оценивания, приведенных в [1], где вероятность правильного распознавания — 0,95.

ЗАКЛЮЧЕНИЕ

Таким образом, в настоящей работе разработаны адаптивные алгоритмы распознавания слов речи на основе модели АРСС. Показано, что автокорреляционные матрицы хранимых эталонов, которые используются для оценок АРСС-параметров выбеливающего фильтра, необходимо пересчитывать с учетом текущих оценок отношения сигнал-шум.

Показано, что для защиты от коррелированной помехи необходима предварительная обработка речевого сигнала цифровым фильтром, построенным на основе модели АРСС. Для предварительной обработки речевого сигнала возможно раздельное оценивание АР- и СС-параметров. Особенностью цифрового фильтра предварительной обработки речевого сигнала на фоне узкополосной помехи и белого шума является то, что до оценивания АР-параметров необходима коррекция автокорреляционной матрицы с учетом дисперсии белого шума.

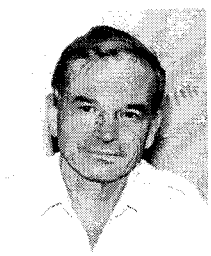
Показано, что для распознавания слов речи могут быть использованы как усредненные меры, так и процедуры динамического программирования, выполняющие нелинейную временную нормализацию.

По найденным рабочим характеристикам проведены экспериментальные исследования алгоритмов распознавания слов речи.

Проведенные исследования алгоритмов распознавания подтверждают возможность получения приемлемого качества распознавания речевых сигналов в условиях действия гауссова белого шума и узкополосных помех и существенного отличия уровней помех на этапе обучения и распознавания.

Литература: 1. *Пресняков И.Н., Омельченко А.В., Омельченко С.В.* Автоматическое распознавание речи в каналах передачи // Радиоэлектроника и информатика. Научно-технический журнал. – 2002. – № 1. – С. 26–31. 2. *Пресняков И.Н., Омельченко С.В.* Автоматическое распознавание раздельных слов и фонем речи // Радиоэлектроника и информатика. Научно-технический журнал. – 2003. – № 4. – С. 41–47. 3. *Пресняков И.Н., Омельченко С.В.* Помехоустойчивые алгоритмы сегментации речи в системах обработки// Радиотехника. Научно-технический сборник. – 2003. – № 131. – С. 165–177. 4. *Пресняков И.Н., Омельченко С.В.* Алгоритмы распознавания фонем речи// Радиотехника. Научно-технический сборник.– 2003. – № 135. – С. 180–189. 5. *Рабинер Л.Р., Шафер Р.В.* Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М. В. Назарова и Ю. Н. Прохорова. – М.: Радио и связь, 1981. – 496 с. 6. Методы автоматического распознавания речи. В 2-х книгах. Книга 1: Пер. с англ. /Под ред. У. Ли. – М.: Мир, 1983. – 328 с. 7. *Марпл.-мл. С.Л.* Цифровой спектральный анализ и его приложения: Пер. с англ. – М.: Мир, 1990. – 584 с. 8. *Дж. Д. Маркел, А.Х. Грей.* Линейное предсказание речи. – М.: Связь, 1980. – 308 с.

Поступила в редколлегию 20.02.2004



Пресняков Игорь Николаевич, доктор технических наук, профессор, заведующий кафедры «Сети связи» ХНУРЭ. Область научных интересов: радиолокация, обработка сигналов, связь.

Омельченко Сергей Васильевич, ассистент кафедры «Сети связи» ХНУРЭ. Область научных интересов: цифровая обработка сигналов, математическое моделирование.