УДК 621.391, 681.327.12

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ РАЗДЕЛЬНЫХ СЛОВ И ФОНЕМ РЕЧИ

ПРЕСНЯКОВ И.Н., ОМЕЛЬЧЕНКО С.В.

Исследуются алгоритмы распознавания фонем и слов речи. Синтез алгоритмов распознавания выполняется с применением различных алгоритмов оценивания признаков и различных мер близости. Исследуется устойчивость алгоритмов распознавания звуковых сигналов к воздействию аддитивного гауссова белого шума и аддитивной гауссовской узкополосной помехи.

Введение

Опираясь на достижения современной лингвистики, вычислительной техники и математической статистики, проводят работу по усовершенствованию алгоритмов распознавания речи, необходимых для решения прикладных задач. Так, для предоставления пользователю мобильной связи дополнительных услуг рационально перейти от клавишного к простому вводу путем побуквенного произнесения слов. Поэтому *актуальными* являются разработки алгоритмов распознавания речи, которые обеспечивают наилучшее соответствие результатов распознавания произнесенным словам и буквам. Система должна быть способна автоматически выявлять и корректировать азбучные (т. е. однобуквенные) аномалии при побуквенном произнесении слов.

Целью исследования является разработка алгоритмов автоматического распознавания слов и фонем речи.

1. Математическая постановка задачи распознавания отдельных слов и фонем речи

Полагается, что на вход системы распознавания поступает временная последовательность отсчетов речевого сигнала $s(n)$, $n = \overline{0, N-1}$, взятых с интервалом дискретизации Δt.

Для создания алгоритмов распознавания важны априорные сведения о вводимых словах и буквах.

Эталоны структурных речевых единиц, включая слова, слоги, буквы (фонемы), 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 выносил бы решения о принадлежности текущих структурных речевых единиц к заданным типам, классам и обеспечивал минимум средней вероятности ошибки распознавания слов, фонем $P_{ош}$ при воздействии аддитивной помехи в канале связи с заданным отношением сигнал-шум q, а также удовлетворял ограничениям на среднюю вероятность ошибки распознавания определенного символа речи.

Вначале для составления хранимых эталонов речевых единиц диктора выполняется сегментация слов, фонем. Подобная сегментация на этапе распознавания речевых единиц позволяет исключить избыточные процедуры принятия решений по сигналам, не несущим речевую информацию либо не являющимся целостными речевыми единицами. Задача сегментации состоит в членении речи на структурные единицы и оценивании их временных границ. Алгоритмы сегментации подробно рассмотрены в [1-5].

2. Алгоритмы распознавания речи

Рассмотрим работу распознавателя изолированных слов и фонем (букв), где выносится решение об определенном слове или фонеме.

Для распознавания речи возможно использование ряда оценок параметров, включая спектральные оценки, измеряемые с помощью набора полосовых фильтров, соответствующих формантным частотам, а также характеристики кодирования на основе линейного предсказания (ЛПК). Такой ряд оценок параметров образован совокупностью измерений в разные моменты времени.

Каждый из приведенных выше наборов признаков обеспечивает хорошее кодирование свойств речи на коротких интервалах времени (отрезках речи), и временные изменения этих характеристик можно, как правило, использовать для описания образа, предназначенного для сравнения с хранимыми эталонами.

Для измерения меры близости образов используется алгоритм, который сравнивает оценки неизвестного испытуемого сигнала с хранящимся эталоном. Выбор меры близости обычно связан с решением следующих двух задач: как выровнять во времени два сигнала разной длительности и как измерить расстояние двух записанных сигналов. Для временного выравнивания известны как простые методы, вроде линейной нормализации во времени, так и сложные, например, динамическое изменение масштаба времени. Для вычисления расстояния используются различные метрики, включая евклидову норму между наборами характеристик, ковариацию взвешенных расстояний, различные спектральные и кепстральные меры и логарифмическое расстояние подобия, определяемое с помощью метода ЛПК. Выбор методов временного

выравнивания и вычисления расстояния зависит также от используемого набора характеристик и допустимого в данной реализации объема вычислений.

После выполнения сегментации слов необходимо принять решение о классе каждого из предъявляемых слов. Алгоритмы распознавания строятся на основе различных мер близости.

Задача распознавания слов и фонем может быть решена с использованием алгоритмов оценивания формантных признаков.

В речевом сигнале, как правило, даже в паузах речи существенно преобладает низкочастотный сигнал, поэтому для увеличения отношения сигнал-шум необходима их коррекция. Кроме того, при блочных алгоритмах обработки наибольший вклад в ошибку оценивания формант будут вносить ее низкочастотные составляющие. Поэтому рационально использовать синхронные методы обработки речи. Значительно повысить устойчивость оценок удается путем предварительной фильтрации речевого сигнала в соответствии с разностным уравнением

$$x_j = s_j - \alpha \cdot s_{j-1}, \qquad (1)$$

где α – коэффициент фильтра.

Экспериментально установлено, что значения α должны выбираться из диапазона 0,8 - 1,0.

В целях получения динамических признаков распознаваемого цифрового сигнала производится разбиение слов на отрезки одинаковой длительности, которая обычно составляет 10-30 мс. При синхронных алгоритмах обработки для вокализованных фрагментов речи длительность отрезков равна периоду основного тона, который несет просодическую информацию и может служить для разметки границ сегментов.

Рассмотрим особенности альтернативной предварительной обработки в условиях действия узкополосных помех.

Полагая, что в пределах выборки речевой сигнал стационарен в широком смысле, алгоритм его выбеливания в частотной области имеет вид

$$\hat{x}(t) = \mathrm{Re}((N)^{-1/2} \sum_{m=0}^{N-1} C(m) H_{кор}(m) \exp(i(2\pi t / N)m)),$$

$$C(m) = (2N)^{-1/2} \sum_{\tau=0}^{2N-1} y_\tau^j \exp(-i(2\pi\tau m / 2N)), \qquad (1a)$$

где $y_i^j = \begin{cases} s_i^j, & i = 0,1,\ldots,(N-1) \\ 0, & i = N,(N+1),\ldots,(2N-1) \end{cases}$ – входные отсчеты; $H(m) = A / \sum_{l \in Z} W(l) \left(S(m+l) \right)^q$ – предварительная оценка амплитудно-частотной характеристики выбеливающего фильтра;

$H_{кор}(m) = H(m)(|d|N - (d-a)m + c)$ – амплитудно-частотная характеристика выбеливающего фильтра;

$d = (\sum_{k=0}^{N-1} k(H(k))^r) / (N \sum_{k=0}^{N-1} (H(k))^r) - 0,5$ – корректирующий параметр;

$S(m) = | N^{-1/2} \sum_{\tau=0}^{N-1} K(\tau) \exp(-i(2\pi m\tau / 2N)) |$ – оценка энергетического спектра;

$K(\tau) = \frac{1}{(N+1-\tau)L} \sum_{j=1}^{L} \sum_{i=0}^{N-\tau} s_{i+\tau}^{(j)} s_i^{(j)}$ – оценка корреляционной функции речевого сигнала.

Экспериментальные исследования речевых сигналов показали, что одномерный в пространстве параметров частот энергетический спектр сигнала в паузе, полученный усреднением 20 выборок по 256 отсчетам, существенно отличается от равномерного, т.е. шум не является белым.

Как показали исследования, использование такого фильтра позволяет существенно повысить отношение сигнал-шум, что обусловливает более высокое качество распознавания речи для ряда рассматриваемых ниже алгоритмов распознавания в условиях действия узкополосных помех.

Далее каждый временной блок (выборка) обрабатывается с использованием временного окна, например, окна Хемминга, в результате чего получается взвешенный отрезок данных $\hat{x}(n)$:

$$\hat{x}(n) = x(n)\hat{w}(n), \qquad (2)$$

где $0 \le n \le N-1$;

$$\hat{w}(n) = 0,54 - 0,46\cos(2\pi n/(N-1)) . \qquad (3)$$

Для распознавания возможно использование спектральных авторегрессионных оценок [1]. Вначале оценивается корреляционная функция и методом Левинсона вычисляются оценки коэффициентов авторегрессии. Затем определяется авторегрессионная спектральная оценка формантных частот в соответствии с выражением

$$\vec{f}_\nu = (F_Д / N) \arg\mathrm{loc}\max(|1 + \sum_{n=1}^{p} \hat{a}[n] \exp(-j2\pi nk)|^{-1}, k = \overline{0,M}), \qquad (4)$$

здесь $M = Z(N/2-1)$, $Z(\ )$ – функция округления числа к целому; $\arg\mathrm{loc}\max(\vec{x})$ – векторная функция, ставящая в соответствие последовательности отсчетов $x_1, x_2, \ldots, x_N$ упорядоченное множество, которое состоит из индексов $f_1, f_2, \ldots, f_L$, удовлетворяющих условию локального максимума: $x_{f_i} > x_{f_{i-1}}$, $x_{f_i} \le x_{f_{i+1}}$.

Рассмотрим особенности формирования формантно-полосных признаков. Согласно этому методу вычисляют спектрально-полосные сигналы, соответствующие вероятному расположению формант, полосы которых

Таблица 1

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

приведены в табл. 1. Граничные частоты $f_в^{(m)}, f_н^{(m)}$ соответствуют m-м формантам при частоте дискретизации 8 кГц .

При этом оценки формантных частот как средних в выделенных полосах вычисляются по формуле

$$\hat{f}^{(m)} = \sum_{i=f_н^{(m)}}^{f_в^{(m)}} i\,|S_i|^2 \Big/ \sum_{i=f_н^{(m)}}^{f_в^{(m)}} |S_i|^2, \qquad (5)$$

где $(f_в^{(m)}, f_н^{(m)})$ – диапазон частот для m-й форманты; S_i – оценка i-й частоты дискретного спектра речевого сигнала.

Аналогично, оценки формантных частот могут вычисляться путем подсчета количества нуль-пересечений речевого сигнала с соответствующего выхода полосового фильтра с заданными граничными частотами $f_в^{(m)}$ и $f_н^{(m)}$, указанными в табл.1 для каждого из блоков (отрезков) речи, которые берутся с 2-х, 3-кратным перекрытием или без него.

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 $\hat{f}_в^{(m)} = \hat{f}^{(m)} + \Delta$, $\hat{f}_н^{(m)} = \hat{f}^{(m)} - \Delta$, где $\hat{f}^{(m)}$ – форманты, вычисленные на предыдущем этапе; Δ – границы диапазона поиска формант. Простейшей среди рекуррентных процедур является двухэтапная.

Относительные амплитуды формант определяют как

$$\hat{A}^{(m)} = \sum_{i=\hat{f}^{(m)}-\Delta f_a}^{\hat{f}^{(m)}+\Delta f_a} |S_i| \Big/ \sum_{m=1}^{4} \sum_{i=\hat{f}^{(m)}-\Delta f_a}^{\hat{f}^{(m)}+\Delta f_a} |S_i|. \qquad (5а)$$

Относительные среднеэффективные амплитуды формант вычисляют как

$$\hat{A}^{(m)} = \frac{\left(\sum_{i=\hat{f}^{(m)}-\Delta f_a}^{\hat{f}^{(m)}+\Delta f_a} \left|S_i^2\right|\right)^{1/2}}{\left(\sum_{m=1}^{4} \sum_{i=\hat{f}^{(m)}-\Delta f_a}^{\hat{f}^{(m)}+\Delta f_a} \left|S_i^2\right|\right)^{1/2}}. \qquad (5б)$$

Определим расстояние как минимальное при всех возможных временных сдвигах $\tau_п$:

$$D_{v,u} = \min_{h=-J,\ldots,J} \frac{\sum_{j=\hat{\tau}_{1,u}}^{\hat{\tau}_{2,u}} D_{v,u}(h,j)}{\hat{\tau}_{2,u} - \hat{\tau}_{1,u} + 1}, \qquad (6)$$

где локальное расстояние может вычисляться как

$$D_{v,u}(h,j) = \sum_{m=1}^{L} |\hat{f}_{j+h-\hat{\tau}_{1,u}+\hat{t}_{1,v}}^{(m)} - \hat{f}_j^{ob(m)}|^r, \qquad (7)$$

либо логарифмическая мера

$$D_{v,u}(h,j) = \log_a\left(\sum_{m=1}^{L} |\hat{f}_{j+h-\hat{\tau}_{1,u}+\hat{t}_{1,v}}^{(m)} - \hat{f}_j^{ob(m)}|\right). \qquad (8)$$

Для меры в пространстве оценок нормированных амплитуд формант

$$D_{v,u}(h,j) = \sum_{m=1}^{L} |\hat{A}_{j+h-\hat{\tau}_{1,u}+\hat{t}_{1,v}}^{(m)} - \hat{A}_j^{ob(m)}|^r. \qquad (8а)$$

Для меры, построенной в пространстве оценок амплитуд и частот формант,

$$\begin{aligned} D_{v,u}(h,j) &= \\ &= \beta \sum_{m=1}^{L} |\hat{A}_{j+h-\hat{\tau}_{1,u}+\hat{t}_{1,v}}^{(m)} - \hat{A}_j^{ob(m)}|^{r1} + \\ &+ \sum_{l=1}^{L} |\hat{f}_{j+h-\hat{\tau}_{1,u}+\hat{t}_{1,v}}^{(m)} - \hat{f}_j^{ob(m)}|^r, \end{aligned} \qquad (8б)$$

где r, r1 – параметры меры (экспериментально получено, что наименьшая вероятность в смысле минимума средней вероятности ошибки Рош распознавания гласных фонем $r \approx 1/2$); $\hat{\tau}_{1,u}, \hat{\tau}_{2,u}$ – оценки временных границ начала и конца u сегмента обучающей выборки; $\hat{t}_{1,v}$, $\hat{t}_{2,v}$ – оценки временных границ начала и конца v-сегмента предъявляемого сигнала.

Номер типа сигнала (вид фонемы, слога или слова) находят в виде

$$i(v) = \arg\min(D_{v,u}, u = \overline{0,L}), \qquad (9)$$

где $\arg\min(f(j), j = \overline{0,L})$ – функция вычисления номера j, при котором функция f(j) минимальна на множестве $j = \overline{0,L}$.

Поиск осуществляется по всем возможным эталонам структурных единиц речи для всех обученных дикторов. С целью улучшить качество распознавания для одной фонемы или слова формируется ряд эталонов для нескольких дикторов.

При структурном распознавании результаты фонемного распознавания используются для принятия решения о конкретном слове.

Расстояние между совокупностью эталонов слова и результатом принятий решений имеет вид

$$S = \sum_{i=1}^{L} f(e_i, P_i) \ge P, \qquad (10)$$

где f(.) – функция сопоставления (при совпадении элемента двоичного слова из заданного алфавита и элемента с двоичным словом, являющимся результатом фонемного распознавания, выносится решение, соответствующее логической единице, в противном случае – логическому нулю).

При этом необходимо обеспечить устойчивость к ошибкам типа пропуск, вставка, перепутывание символа. Поэтому возможно использование мно-

жества искаженных эталонов, а также динамические методы сопоставления на уровне перехода от фонем к словам.

3. Экспериментальные исследования алгоритмов распознавания структурных единиц речи

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

Оценки траекторий формантных признаков были получены с использованием различных алгоритмов их оценивания (4),(5,б).

Из сравнений рис. 1, а и б видно, что траектории оценок формантных частот буквы "ю" по подсчету числа нуль-пересечений сигналов с выходов фильтров (а) соответствуют результатам, полученным по методу периодограмм (б).

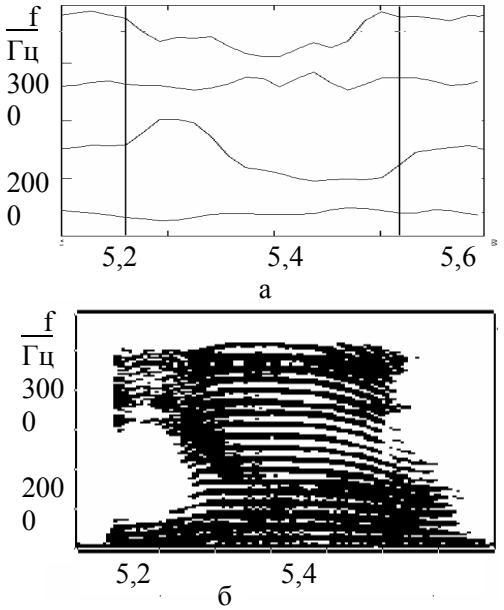

Рис. 1. Траектории оценки формант буквы «ю»

На рис. 2 показана динамика изменений оценок формант сигналов для слов, вычисленных в соответствии с алгоритмом (7) для r=1.

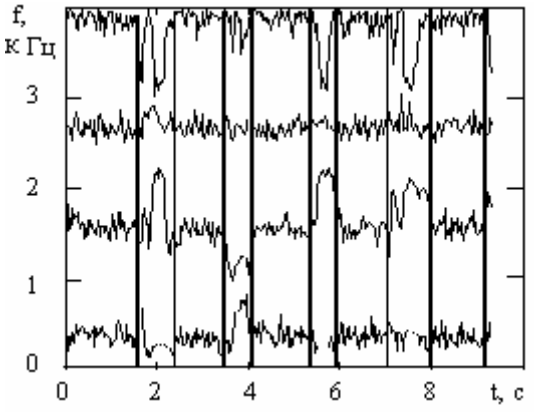

Рис. 2. Траектории оценки формант четырех слов

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

Таблица 2

Признак	РП	$\hat{P}_{прав.ср.}$
AR	ЛМ	0,95
ЧАНСП2	М1/2	0,98
АНСП2	М1/2	0,6
ЧНСП2	М1/2	0,97
ЧНСП	М1/2	0,9
ЧНСП	ЛОМ	0,82
ЧНСП	ЛМ	0,95
ЧНСП	КМ	0,82

В оценочных тестах система распознавания правильно находила слово в среднем в 95-98 % попыток, несмотря на то, что акустический распознаватель правильно опознавал буквы приблизительно в 60-92 % случаев. В табл.2 приведены результаты исследования вариантов устройств распознавания слов, отличающихся либо типом алгоритма оценивания признаков, либо типом решающих функций.

Варианты алгоритмов вычисления признаков для распознавания слов: AR – по предсказанию; ЧАНСП2 – 2-этапное определение количества нулей в полосах формантных частот и нормированных амплитуд формант; АНСП2 – 2-этапное определение нормированных амплитуд формант; ЧНСП2 – 2-этапное определение количества нулей в полосах формантных частот; ЧНСП – количество нулей в полосах формантных частот.

Варианты решающих правил (РП) – мер близости признаков: ЛМ – линейная мера; М1/2 – мера степени 1/2; КМ – квадратичная мера; ЛОМ – логарифмическая мера.

В табл.3 приведены результаты исследования вероятности принятия решений о наличии в выделен-

Таблица 3

р	и	ы	о	ю	я	е	ё	у	а	э	ї
и	0,82	0,36	0	0	0	0,09	0	0	0	0	0
ы	0	0,64	0	0	0	0,09	0	0	0	0,09	0
о	0	0	1	0	0	0	0	0	0	0	0
ю	0	0	0	0,91	0	0	0,09	0	0	0	0
я	0	0	0	0	1	0	0	0	0	0	0
е	0	0	0	0	0	0,82	0	0	0	0	0,18
ё	0	0	0	0,09	0	0	0,91	0	0	0	0
у	0	0	0	0	0	0	0	1	0	0	0
а	0	0	0	0	0	0	0	0	1	0	0
э	0	0	0	0	0	0	0	0	0	0,91	0
ї	0,18	0	0	0	0	0	0	0	0	0	0,82

Таблица 4

р	и	ы	о	ю	я	е	ё	у	а	э	ї
и	0,82	0,64	0	0	0	0	0	0	0	0,18	0
ы	0	0,36	0	0	0	0,09	0	0	0	0	0
о	0	0	1	0	0	0	0	0	0	0	0
ю	0	0	0	1	0	0	0,09	0	0	0	0
я	0	0	0	0	1	0	0	0	0	0	0
е	0	0	0	0	0	0,82	0	0	0	0	0,09
ё	0	0	0	0	0	0	0,91	0	0	0	0
у	0	0	0	0	0	0	0	1	0	0	0
а	0	0	0	0	0	0	0	0	1	0	0
э	0	0	0	0	0	0	0	0	0	0,82	0
ї	0,18	0	0	0	0	0,09	0	0	0	0	0,91

ных сегментах русско-украинских гласных букв “и, ы, о, ю, я, е, ё, у, а, э, ї” для алгоритма с одноэтапным определением формантных частот по нуль-пересечениям сигнала и растоянием (6), (7) с параметром r=×, а в табл.4 – для растояния (6), (8) с логарифмической мерой. Буквы ю, я, е, ё, ї являются дифтонгами, которые начинаются с й и затем постепенно переходят в у, а, э, о, и.

Из полученных результатов экспериментальных исследований при фонемном (побуквенном) распознавании можно сделать вывод, что наилучшими в смысле минимума средней вероятности ошибки по всем символам при низком уровне аддитивных помех являются алгоритмы с двухэтапным определением формантных частот по нуль-пересечениям сигнала для расстояния (6), (7) с r=1/2 и расстояния (6), (8) с логарифмической мерой. Однако не все символы распознаются с равным качеством. Так, для алгоритма (8) максимальная ошибка на один символ для буквы ы будет 0,64, а для алгоритма (7) с параметром r=1/2 – 0,36.

При ограничении допустимой максимальной ошибки, приходящейся на один символ «ы», оптимальным будет алгоритм с r=1/2.

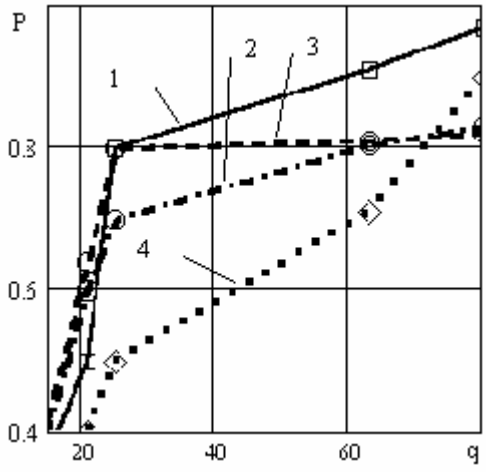

Рис.3. Зависимость правильного распознавания слов от отношения среднеквадратического значения наблюдаемого сигнала к среднеквадратическому отклонению аддитивной помехи

Методом статистических испытаний получены зависимости оценки вероятностей правильного распознавания слов от отношения сигнал-шум q. Испытания проводились на реальных выборках звуковых сигналов, введенных в ЭВМ с выхода микрофона.

На рис.3 приведены зависимости вероятности правильного распознавания слов от отношения среднеквадратического значения наблюдаемого сигнала на фоне естественного шума к среднеквадратическому отклонению дополнительно сгенерированной на ЭВМ аддитивной помехи типа гауссов белый шум при одноэтапном оценивании формантных частот путем счета числа нуль-пересечений с выходов полосовых фильтров для: 1 – решающей функции (7) с линейной мерой; 2 – решающей функции (8) с логарифмической мерой; 3 – решающей функции (7) с квадратичной мерой; 4 – решающей функции (7) при параметре r =1/2.

Экспериментальные исследования спектрально-полосных алгоритмов распознавания слов речи с одноэтапным определением количества нулей в полосах формант проводились методом статистических испытаний на выборках 10-и сигналов для каждого из 3-х различных дикторов. По выборкам оценивались параметры решающего правила, а контрольные выборки реальных сигналов использовались для оценивания качества распознавания сигналов.

Наибольшая вероятность правильного распознавания слов получена для алгоритма с оценкой формант по количеству нуль-пересечений в полосах формантных частот и с линейной мерой при отношении сигнал-шум q>20. Сравнивая кривые рис.4 и результаты, полученные в [1], можно сделать вывод, что вероятности правильного принятия решения для алгоритма с оценкой формант по количеству нуль-пересечений в полосах формантных частот и с мерой (6), (7) при r=1/2 в условиях отсутствия дополнительных белых шумов не хуже, чем для алгоритмов, построенных в пространстве авторегрессионных (АР) спектральных оценок, а в условиях действия белого гауссова шума более устойчив алгоритм АР спектральных оценок.

В табл. 5 приведены средние вероятности правильного распознавания $\hat{P}_{\text{п.ср.}}$ гласных фонем (букв) для алгоритмов вычисления признаков (Пр):

AR – по предсказанию; ЧАНСП2 – 2-этапное определение количества нулей в полосах формантных частот и нормированных амплитуд формант для $\Delta = 10$; ЧНСП2 – 2-этапное определение количества нулей в полосах формантных частот для $\Delta = 10$; ЧНСП – количество нулей в полосах формантных частот; ЧССП – средняя частота в полосах формантных частот, а также для различных РП - мер близости признаков: ЛМ – линейная мера; М1/2 – мера степени 1/2; КМ – квадратичная мера; ЛОМ- логарифмическая мера.

В целях изучения совместного вклада четырех оценок формантных частот на результат правильного принятия решений в эксперименте из алгоритмов удалялся ряд формант. Исследовались средние вероятности правильного распознавания десяти слов $\hat{P}_{\text{прав. ср.}}$ от номеров используемых формант для алгоритмов с двухэтапным определением количества нулей в полосах формантных частот и с параметром меры (6), (7) r=1/2. Экспериментально получено, что наибольший вклад в процесс распознавания вносит 2-я форманта, при этом вероятность правильного распознавания

Таблица 5

Пр	РП	$\hat{P}_{\text{п.ср.}}$
AR	ЛМ	0,6
ЧНСП2	М1/2	0,91
ЧНСП2	ЛОМ	0,92
ЧНСП2	ЛМ	0,89
ЧНСП	М1/2	0,87
ЧНСП	ЛОМ	0,87
ЧНСП	ЛМ	0,81
ЧНСП	КМ	0,73
ЧССП	ЛОМ	0,73
ЧССП	М1/2	0,73
ЧССП	ЛМ	0,73

$\hat{P}_{прав.ср.}$=0,8. В случае применения в алгоритме второй форманты наибольший вклад в процесс правильного принятия решения дает третья форманта, при этом $\hat{P}_{прав.ср.}$=0,9. При использовании второй и третьей форманты качество распознавания улучшается при дополнительном использовании первой форманты – $\hat{P}_{прав.ср.}$=0,95. Средняя вероятность правильного распознавания тем же методом по частотам первой форманты $\hat{P}_{прав.ср.}$=0,5; третей форманты – $\hat{P}_{прав.ср.}$=0,4; четвертой форманты – $\hat{P}_{прав.ср.}$=0,1. В то же время их совместное использование наиболее эффективно для повышения средней вероятности правильного распознавания в случае первой, третьей формант – $\hat{P}_{прав.ср.}$=0,6, а в случае третьей и четвертой формант – $\hat{P}_{прав.ср.}$=0,5.

В целях определения характеристик и проверки работоспособности алгоритма предварительной обработки в условиях действия узкополосных случайных процессов проводилось математическое моделирование помех на ЭВМ следующим образом. Сигналы помех генерировались в виде амплитудно модулированных сигналов

$$n_j = A(1+\xi_j)\cos(2\pi jF/F_д), \qquad (11)$$

где $F_Д$ – частота дискретизации.

Модулирующий сигнал ξ_j удовлетворяет уравнению авторегрессии первого порядка

$$\xi_j = \alpha\xi_{j-1} + (1-\alpha)\eta_j. \qquad (12)$$

В эксперименте для моделируемых узкополосных помех задавался коэффициент авторегрессии $\alpha = 0{,}99$.

Порождающий процесс η является гауссовским с математическим ожиданием m=0 и среднеквадратическим отклонением $\sigma = 4$. Длительность наблюдаемых реализаций модельных сигналов принималась равной длительности распознаваемого речевого сигнала. При проведении статистического эксперимента задавалась центральная частота узкополосной помехи F=560Гц.

Распознавание по наблюдаемой аддитивной смеси сигнала s_j и помехи n_j производилось в соответствии с алгоритмом (6) 2-этапного определения количества нулей в полосах формантных частот $\Delta = 10$ с параметром меры (6), (7) r=1/2. Для второго этапа параметр $\Delta = 10$ соответствует оценкам нижних граничных частот формантно-полосных фильтров в виде $f_н = f_i$ -156 Гц, а оценкам верхних граничных частот - $f_в = f_i$ +156 Гц, где f_i – оценка i-й формантной частоты на первом этапе.

Для заданных условий эксперимента и отношений сигнал-шум по мощности q^2 получены значения оценок средних вероятностей правильного распознавания слов $\hat{P}_{п.1}$ алгоритмов с первым (1) и $\hat{P}_{п.1а}$ вторым (1а) видом предварительной обработки, которые приведены в табл. 6.

Результаты экспериментальных оценок средней вероятности правильного распознавания слов $\hat{P}_{п.1}$ алгоритмов предварительной обработки (1) и вероятности правильного распознавания слов $\hat{P}_{п.1а}$ алгоритмов предварительной обработки (1а) для ряда значений центральной частоты F узкополосной помехи приведены в табл. 7. Из полученных результатов видно, что без предварительной обработки (1а) алгоритм наиболее чувствителен к воздействию узкополосной случайной помехи в частотных полосах 2-й и 1-й формант.

Таблица 6

q^2	$\hat{P}_{п.1}$	$\hat{P}_{п.1а}$
0.21	0.4	0.8
1	0,7	0,9
2,4	0,7	0,9
21,2	0.8	0,9

Таблица 7

F, Гц	$\hat{P}_{п.1}$	$\hat{P}_{п.1а}$
560	0,7	0,9
1500	0,5	0,9
2700	0,8	0,9
3500	0,8	0,95

На рис.4,а приведена усредненная по 20 выборкам корреляционная функция аддитивной смеси речевого сигнала и узкополосной помехи с центральной частотой 1500 Гц и соотношением сигнал-шум по мощности q^2=1 до фильтрации, а на рис.4,б – корреляционная функция речевого сигнала в паузе после выбеливания аддитивной смеси речевого сигнала и помехи алгоритмом предварительной обработки (1а).

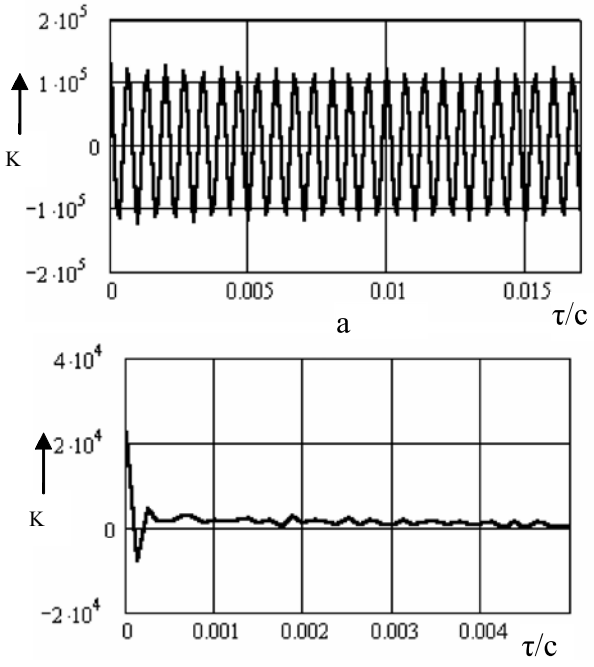

Рис.4. Усредненная оценка корреляционной функции: а – для речевого сигнала со случайной узкополосной помехой с центральной частотой F=1,5 кГц; б – после выбеливания с алгоритмом (1а)

На рис.5 приведена оценка амплитудно-частотной характеристики выбеливающего фильтра с передаточной характеристикой H(m).

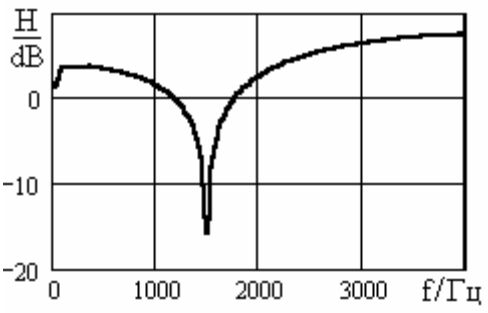

Рис.5. Оценка амплитудно-частотной характеристики выбеливающего фильтра (1а)

На рис.6,а приведен энергетический спектр сигнала с помехой в паузе, а на рис. 6, б – энергетический спектр речевого сигнала в паузе, полученный в результате выбеливания сигнала фильтром (1а).

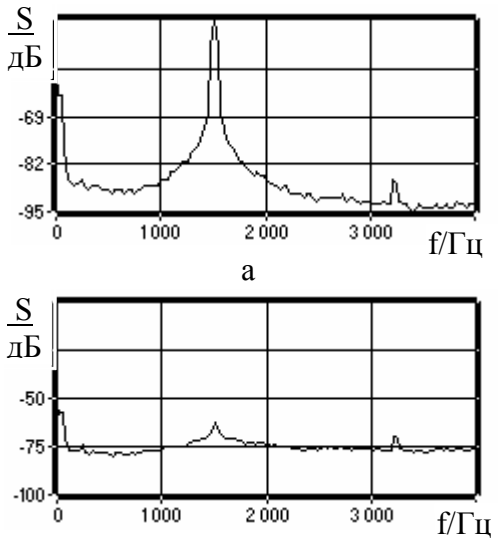

Рис.6. Энергетический спектр: а – для речевого сигнала со случайной узкополосной помехой с центральной частотой F=1,5 кГц; б – для результата выбеливания алгоритмом (1а)

Траектории оценок формантных частот четырех слов при действии узкополосной помехи с соотношением сигнал-шум по мощности $q^2 = 1$ и частотой 1500 Гц приведена на рис. 7, а после обработки алгоритмом (1а) с параметрами c=0, a=0,25 – на рис. 8. Из рис. 7 видно, что оценка траектории второй формантной частоты ухудшается из-за подавления речевого сигнала помехой, а на рис. 8 после обработки (1а) – восстановлена и подобна оценке траектории второй формантной частоты сигнала без помех, показанной на рис. 2.

Проведенные исследования подтверждают эффективность алгоритма предварительной обработки (1а) в условиях действия узкополосных случайных помех.

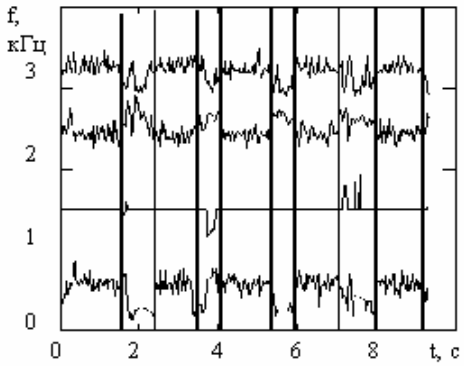

Рис. 7. Траектории оценки формант четырех слов (узкополосная помеха с частотой 1500 Гц)

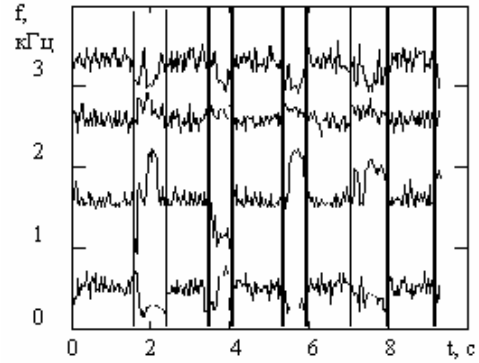

Рис.8. Траектории оценки формант четырех слов после фильтрации

Выводы

Таким образом, в настоящем исследовании разработаны алгоритмы распознавания слов и фонем (букв) речи для разных мер близости. По найденным рабочим характеристикам проведены сравнительные исследования алгоритмов распознавания букв и слов речи в телекоммуникации для различных видов решающих функций и разных оценок формантных частот. Проведенные исследования алгоритмов распознавания подтверждают возможность получения приемлемого качества распознавания речевых сигналов по формантным признакам в условиях действия гаусова белого шума и узкополосных помех.

Литература: 1. *Пресняков И.Н., Омельченко А.В., Омельченко С.В.* Автоматическое распознавание речи в каналах передачи // Радиоэлектроника и информатика. 2002. № 1. С. 26-31. **2.** *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. 496с. **3.** *Методы* автоматического распознавания речи: В двух книгах. Пер. с англ. /Под ред. У. Ли. М.: Мир, 1983. Кн. 1. 328с. **4.** *Марпл.-мл. С.Л.* Цифровой спектральный анализ и его приложения: Пер. с англ. М.: Мир, 1990. 584с. **5.** *Маркел Дж. Д., Грей А. Х.* Линейное предсказание речи. М.: Связь, 1980. 308с.

Поступила в редколлегию 04.04.2003

Рецензент: д-р техн. наук, проф. Руденко О.Г.

Пресняков Игорь Николаевич, д-р техн. наук, профессор, зав. кафедрой “Сети связи” ХНУРЭ. Адрес: Украина, 61000, Харьков, пр. Победы, 54-б, кв. 44, тел. 702-14-29.

Омельченко Сергей Васильевич, ассистент кафедры “Сети связи” ХНУРЭ. Адрес: Украина, 61000, Харьков, ул. Кузнецкая, 102а, тел. 702-14-29.