УДК 621.391
DOI: 10.15587/2313-8416.2018.129703

АЛГОРИТМЫ СЕГМЕНТАЦИИ РЕЧЕВОГО СИГНАЛА НА ФОНЕ КОРРЕЛИРОВАННОЙ ПОМЕХИ

© С. В. Омельченко

В статье рассмотрены алгоритмы сегментации на основе оценок формант и антиформант. Получен алгоритм сегментации речи с использованием моментных функцій третего и четвертого порядка. Предложено с целью подавления коррелированных помех использовать цифровую фильтрацию на основе модели авторегрессии скользящего-среднего. Получены оценки дисперсии оценивания временных границ слов для ряда предложенных алгоритмов сегментации речи

Ключевые слова: *сегментация речи, модель авторегрессии скользящего-среднего, моментные функции, форманти, фонемы, коррелированные помехи*

1. Введение

Под сегментацией речи обычно понимают расчленение речевого потока на некоторые элементы – фонемы, слоги, слова (при распознавании слитной речи), как правило, связанные с фонетическим представлением речевых сообщений. Существующие методы автоматической сегментации речи плохо защешены от воздействия помех и плохо адоптируютя к измениниям окружающей обстановки. Поэтому задача поиска помехоустойчивых методов сегментации является актуальной задачей.

2. Литературный обзор

Для создания алгоритмов распознавания речи, устойчивых к действию помех, необходима высокая точность оценок временных границ сегментов речи в условиях действия помех. В качестве информативных параметров, используемых для сегментации, могут быть различные характеристики речевых сигналов. К ним относятся частота основного тона [1, 2] формантные частоты [2], признак вокализованности [2], мощность сигнала в разных полосах частот сигнала [2], длительности произносимых фонем [1, 2] сегментация по корреляции между равноотстоящими спектрами [2–4]. Однако необходимы дальнейшее исследования алгоритмов сегментации речи устойчивых к действию коррелированных помех.

3. Цель и задачи исследования

Цель исследования – разработка алгоритмов сегментации речи устойчивых к действию помех в канале святи.

Для достижения цели были поставлены следующие задачи:

1. Рассмотреть возможность подавления коррелированных помех.

2. Розработать методы сегментации речи, которые являются устойчивыми к действию помех, характерных для речевого канала и телефонных каналов связи.

3. Провести эксперементальные исследования разработанных алгоритмов.

4. Материалы и методы исследования

Рассмотрим математическую постановку задачи сегментации речи и основные особенности её решения.

Априорная информация в виде эталонов сигнала, необходимая для алгоритмов распознавания, задаётся в виде классифицированных обучающих выборок в паузах между словами для каждого из дикторов. Считается, что время появления слова в слитном речевом сигнале априори неизвестно и заданы ограничения на длительность пауз между слогами слов.

Качество $\vec{K}$ алгоритма s будем оценивать величиной дисперсии D(s) оценки временного положения сегментов при отсутствии внешней аддитивной помехи и устойчивостью $k_{уст}(s)$ алгоритма s к воздействию аддитивной помехи

$$\vec{K}(s) = (D(s), k_{уст}(s)). \quad (1)$$

Под показателем устойчивости $k_{уст}(s)$ понимается дисперсия оценки временного положения сегментов при воздействии аддитивной помехи в канале с заданным отношением сигнал-шум [1, 2].

Необходимо построить оптимальный алгоритм определения по реализациям речи моментов времени начала и конца слов, который обеспечивает максимум целевой функции в классе робастных алгоритмов.

5. Предварительная обработка речевого сигнала

Рассмотрим предварительную обработку речевого сигнала цифровым фильтром, построенным на основе модели авторегрессии скользящего-среднего (АРСС) [5]. Такой фильтр необходим для исключения коррелированных помех из сигнала и выравнивания АЧХ распознаваемых сигналов [6–11]. Полагается, что априори известен интервал времени, в течении которого отсутствует речь (пауза). Такой интервал времени используется для оценивания АРСС-параметров фильтра предварительной обработки.

Для оценивания АРСС- параметров, как правило, применяются процедуры раздельного оценива-

ния параметров авторегрессии (АР) и параметров скользящего-среднего (СС) [5]. 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.

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

$$n_t = \sum_{u=1}^{p} a_u n_{t-u} + \xi_t, \qquad (2)$$

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:

$$[r]\cdot\bar{a} = \bar{r}, \qquad (3)$$

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 $r_j = R_j / R_0$, где

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

– 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

$$y_t = x_{yt} + n_{\delta t} \qquad (4)$$

с корреляционной функцией

$$R_{yj} = R_{nyj} + D_n \cdot \delta(j), \qquad (5)$$

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

$$R_{nyj} = R_{yj} - D_n \cdot \delta(j), \qquad (6)$$

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

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

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

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

Вектор оценок коэффициентов АР находится из выражения

$$\bar{a} = [r]^{-1} \cdot \bar{r}. \qquad (7)$$

Алгоритм оценивания ошибки предсказания описывается выражением

$$y_t = x_t - \sum_{u=1}^{p} \hat{a}_u x_{t-u}, \qquad (8)$$

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

Оценка нормированной корреляционной функции ошибки предсказания сигнала в паузе

$$K_{yj} = \frac{1}{(T+1-j)\cdot(L2+1-L1)} \sum_{v=L1}^{L2} \sum_{i=0}^{T-j} (y_{i+j}^{(v)} \cdot y_i^{(v)}), \qquad (9)$$

где v-номер выборки, T-период наблюдения.

Фильтрация сигнала ошибки предсказания описывается разностным уравнением

$$s_t = -\sum_{u=1}^{p} b_u s_{t-u} + y_t, \qquad (10)$$

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

Нормированная АЧХ фильтра

$$H(n2\pi/T) = \frac{|1 - \sum_{k=1}^{p} (a_k \cdot e^{-ikn2\pi/T})|}{|\sum_{k=0}^{p} (b_k \cdot e^{-ikn2\pi/T})|}. \qquad (11)$$

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.

6. Алгоритмы сегментации речи по энергетическим признакам

Рассмотрим алгоритмы сегментации речи по энергетическим признакам.

В результате применения декорелирующего фильтра, алгоритмы обнаружения могут быть упрощены за счет декорреляции временных отсчетов речевого сигнала [2].

При обеспечении некоррелированности признаков и равенства дисперсий в координатном представлении в алгоритме обнаружения речевого сигнала по энергетическим признакам выносится решение о наличии речевой информации в к-ой выборке, если выполняется неравенство

$$H_1 : (l(k)) > \Lambda, \quad (12)$$

где $l(k) = \sum_{i=1}^{N} \left| S_i^k \right|^2$, а S_i^k -l-ый отсчет k-ой выборки речевого сигнала.

В противном случае выносится решение о наличии паузы.

Порог Λ в общем случае вычисляется как

$$\Lambda = \frac{2\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2} \ln\left[\left(\frac{\sigma_1^2}{\sigma_0^2}\right)^n c\right] > 0, \quad \sigma_1 > \sigma_0. \quad (13)$$

Для критерия Неймана-Пирсона при заданном α порог преобразуется к виду $\Lambda = \sigma_0^2 \chi_\alpha^2$, где χ_α^2 - выраженное в процентах отклонение случайной величины, распределенной по закону χ^2 с n степенями свободы.

Вычисление пороговых уровней Λ, также может производится экспериментально по результатам определения локальных минимумов близлежащих справа или слева (в зависимости от задачи) от глобального максимума гистограммы распределения решающей статистики [2].

7. Алгоритмы сегментации речи в пространстве оценок моментных функций

Моментные функции третьего порядков стационарного процесса определяются выражениями

$$m_2[i, j] = m_2[0, j - i]. \quad (14)$$

Трёхмерные моментные функции стационарного процесса определяются выражениями

$$m_3[i, j, k] = m_3[0, j - i, k - i]. \quad (15)$$

Четырёхмерные моментные функции стационарного процесса определяются выражениями

$$m_4[i, j, k, n] = m_3[0, j - i, k - i, n - i]. \quad (16)$$

Выборочные значения оценок моментных функции определяются выражением

$$m_3[0, j - i, k - i] = \\ = \frac{1}{N - h} \sum_{t=1}^{N-h} x[t] \cdot x[t + i] \cdot x[t + k], \quad (17)$$

$$m_4[0, j - i, k - i, n - i] = \\ = \frac{1}{N - h} \sum_{t=1}^{N-h} x[t] \cdot x[t + i] \cdot x[t + k] \cdot x[t + n], \quad (18)$$

где h – максимальное значение сдвига для каждого набора (j-i), (k-i).

Решение на основе оценок трёхмерных моментных функций принимается в соответствии с выражением

$$H_1 : R(k) < \Lambda, \quad (19)$$

где $\mathrm{R(k)} = \sum_{u=0}^{p1} \sum_{v=0}^{p2} \mathrm{sgn}\left(m_{3,k}(0,u,v)\right) \oplus \mathrm{sgn}\left(m_3^T(0,u,v)\right)$,

где функция $\mathrm{sgn}(x) = \begin{cases} 1, x \geq 0; \\ 0, x < 0. \end{cases}$

Вычисление пороговых уровней Λ, производится экспериментально по результатам определения локальных минимумов близлежащих справа или слева от глобального максимума гистограммы распределения решающей статистики.

Решающее правило на основе оценок знаковых функций моментной функции может быть представлено в виде

$$H_1 : R(k) < \Lambda, \quad (20)$$

где $\mathrm{R(k)} = \sum_{u=0}^{p1} \sum_{v=00}^{p2} (sign\left(m_{3,k}(0,u,v)\right) \cdot sign\left(m_3^T(0,u,v)\right))$,

$$sign(x) = \begin{cases} 1, x \geq 0; \\ -1, x < 0. \end{cases}$$

Решение на основе оценок моментной функции принимается в соответствии с выражением

$$H_1 : R(k) < \Lambda, \quad (21)$$

где среднее расстояние можно вычислить в виде

$$\mathrm{R(k)} = \sum_{u=0}^{p1} \sum_{v=0}^{p2} (m_{3,k}(0,u,v)) - m_3^T(0,u,v))^2.$$

Расстояния трёхмерных моментных функций может быть представлено в виде

$$\mathrm{R(k)} = \sum_{u=0}^{p1} \sum_{v=0}^{p2} -(\left|m_{3,k}(0,u,v)) - m_3^T(0,u,v)\right|^r + \alpha)^{-1},$$

$$R(k) = \sum_{u=0}^{p1}\sum_{v=0}^{p2}\sum_{n=0}^{p} -(\left|m_{3,k}(0,u,v,n)) - m_3^{T}(0,u,v,n)\right|^{r} + \alpha)^{-1},$$

где r – параметр расстояния.

8. Алгоритм сегментации речи по совокупности формант и антиформант

Для сегментации возможно использование оценок формант и антиформант. Авторегрессионые спектральные оценки формантных частот вычисляются в соответствии с выражением

$$\vec{f}_{v} = \frac{F_o}{N}\arg loc\max_{k}\{|\frac{\sum_{n=1}^{p}\hat{b}(n)\exp(-j2\pi nk)}{1-\sum_{n=1}^{p}\hat{a}(n)\cdot\exp(-j2\pi nk)}|,$$

$$k=\overline{0,M}\},$$

где $\vec{f} = \arg loc\max(\vec{x})$ – векторная функция, задающая соответствие элементам входной последовательности $x_1,x_2,\ldots,x_N$ элементам выходной последовательности упорядоченное множество номеров локальных максимумов $\{ f_i, i=\overline{0,L}\}$; вектор оценок $\vec{f}_v = \{ f_{i,v}, i=\overline{0,L}\}$, L -количество локальных максимумов в спектре; $F_o = 1/\Delta t$ – частота дискретизации сигнала, Δt - период дискретизации сигнала; $M = z[N/2-1]$; Z[y] – функция округления к ближайшему целому числу.

Авторегрессионые спектральные оценки частот антиформант вычисляются в соответствии с выражением

$$\vec{f}_{a,v} = \frac{F_o}{N}\arg loc\min_{k}\{|\frac{\sum_{n=1}^{p}\hat{b}(n)\exp(-j2\pi nk)}{1-\sum_{n=1}^{p}\hat{a}(n)\cdot\exp(-j2\pi nk)}|,$$

$$k=\overline{0,M}\},$$

где $\vec{f} = \arg locmin(\vec{x})$ – векторная функция, задающая соответствие элементам входной последовательности $x_1,x_2,\ldots,x_N$ элементам выходной последовательности упорядоченное множество номеров локальных минимумов $\{ f_i, i=\overline{0,L}\}$; вектор оценок $\vec{f}_v = \{ f_{i,v}, i=\overline{0,L_{мин}}\}$, Lмин -количество локальных максимумов в спектре.

После выполнения сегментации фонем необходимо принять решение о наибольшей степени близости в пространстве признаков произносимой фонемы и одной из фонем обучающих выборок.

Решение о начале нового сегмента фонем в очередной выборке принимается по результату сравнения с порогом значений $R_n^{фон}$, вычисленных по формуле

$$R_n^{фон} > \Lambda,$$

где $R_n^{фон}$ – функционалы, построенные на основе метрик в пространстве L_1, L_2

$$R_n^{фон} = \sum_{i=1}^{L}\min_{j\in[-J,J]}\alpha_{i,j}^{l}\cdot|\hat{f}_i(n) - \hat{f}_{i+j}^{П}(n+1)|^{q} +$$

$$+\sum_{i=1}^{La}\min_{j\in[-J,J]}\alpha_{i,j}^{l}\cdot|\hat{fa}_i(n) - \hat{fa}_{i+j}^{П}(n+1)|^{q},$$

где $\hat{f}_i(n)$, $\hat{fa}_i(n)$ – оценки частот i-ой форманты n-го сегмента; $\alpha_{i,j}^{l}$ – весовые коэффициенты, $i=\overline{-J,J}$; $j=\overline{-J,J}$; q принимает значения 1 или 2 в зависимости от вида критерия близости.

На основе первичной сегментации слов по формантным признакам выносится решение о наличии речевой информации в n-ом сегменте в случае если

$$H_1 : R_n^{слов} < \Lambda,$$

$$R_n^{слов} = \sum_{i=1}^{L(n)}\min_{j\in[-J,J]}\alpha_{i,j}^{l}\cdot|\hat{f}_i(n) - \hat{f}_{i+j}^{П}|^{q} +$$

$$+\sum_{i=1}^{L(n)}\min_{j\in[-J,J]}\alpha_{i,j}^{l}\cdot|\hat{fa}_i(n) - \hat{fa}_{i+j}^{П}|^{q},$$

где $\hat{f}_i(n)$, $\hat{fa}_i(n)$ – оценки частот i-ой форманты и антиформанты n-го сегмента; $\hat{fa}_{i+j}^{П}$, $\hat{fa}_{i+j}^{П}$ - эталонные оценки частот i-ой форманты и антиформанты, полученные усреднением оценок для нескольких сегментов, соответствующих паузе речи; $\alpha_{i,j}^{l}$, $i=\overline{-J,J}$; $j=\overline{-J,J}$; q-весовые коэффициенты.

9. Результаты экспериментального исследования алгоритмов сегментации речи и их обсуждение

Исследования описанных выше методов сегментации выполнены по выборкам реальных речевых сигналов для разных дикторов. Оценивание показателей качества производилось для алгоритмов выполняющих сегментация слов речи по энергетическим признакам с выбеливанием и без выбеливания, по признакам формант и антиформант для порядка модели 12.

С целью звукового контроля качества сегментации речи с помощью экспертов проведены экспериментальные исследования. По отсчетам звукового сигнала, следующих в результате дискретизации с частотой 8 кГц, проводились оценки временных границ начала и конца каждого из 10 слов речи.

В табл. 1 приведены результаты исследования 3 варианта устройств сегментации слов, отличающихся типом алгоритма оценивания начала и конца слова. Из таблицы видно, что тип устройства в смысле критерия максимума дисперсии оценивания временного положения слов, зависит от требований устойчивости.

Если задать допустимое значение показателя устойчивости $K_{уст}(s)$ соответствующего отношению сигнал шум q=13, то наилучшим по показателю дисперсии оценивания временного положения D будет алгоритм сегментации слов по энергетическим признакам с выбеливанием.

Таблица 1

Результаты исследований устройств сегментации слов

Алгоритмы сегментации слов	D, $с^2$	D,$с^2$ при q=13
По энергетическим признакам с выбеливанием	0,00023	0,0005
По энергетическим признакам без выбеливания	0,0022	0,0035
Формант и антиформант для порядка модели 12	0,0019	0,0020

10. Выводы

1. В статье рассмотрены возможности подавления коррелированных помех за счет использования цифрового фильтра на основе модели авторегрессии скользящего-среднего.

2. Получены алгоритмы оценивания временных границ слов речи на основе моментных функций, формантных и антиформантных признаков с выбеливанием. Рассмотрены различные пути решения сформулированной задачи сегментации речевых сигналов.

3. На основе эксперименалных исследований показана эфективность предложенных алгоритмов оценивания временных границ слов речи на основе ряда энергетических признаков, формантных и антиформантных признаков с выбеливанием.

Литература

1. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов / под ред. М. В. Назарова, Ю. Н. Прохорова. М.: Радио и связь, 1981. 496 с.

2. Пресняков И. Н., Омельченко С. В. Помехоустойчивые алгоритмы сегментации речи в системах обработки // Радиотехника. 2003. № 131. С. 165–177.

3 Сорокин В. Н., Цыплихин А. И. Сегментация и распознавание гласных // Информационные процессы. 2004. Т. 4, № 2. С. 202–220.

4. Сорокин В. Н., Цыплихин А. И. Сегментация речи на кардинальные элементы // Информационные процессы. 2006. Т. 6, № 3. С. 177–207.

5. Марпл С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 584 с.

6. Пресняков И. Н., Омельченко С. В. Автоматическое распознавание раздельных слов и фонем речи // Радиоэлектроника и информатика. 2003. № 2. С. 41–47.

7. Пресняков И. Н., Омельченко С. В. Алгоритмы распознавания фонем речи // Радиотехника. 2003. № 135. С. 180–189.

8. Пресняков И. Н., Омельченко С. В. Распознавание речевого сигнала на фоне коррелированной помехи // Радиотехника. 2004. Вып. 137. С. 23–30.

9. Пресняков И. Н., Омельченко С. В. Алгоритмы распознавания речи // Автоматизированные системы управления и приборы автоматики. 2004. № 126. С. 136–145.

10. Пресняков И. Н., Омельченко С. В. Распознавание фонем речи // Радиоэлектроника и информатика. 2004. № 3. С. 59–63.

11. Пресняков И. Н., Омельченко С. В. Распознавание речевого сигнала на фоне белого шума и узкополосной помехи // Прикладная радиоэлектроника. 2004. Т. 3, № 2. С. 29–35.

Рекомендовано до публікації д-р техн. наук Безрук В. М.
Дата надходження рукопису 06.03.2018

Омельченко Сергей Васильевич, кандидат технических наук, доцент, кафедра информационно сетевая инженерия, Харьковский национальный университет радиоэлектроники, пр. Науки, 14, г. Харьков, Украина, 61166
E-mail: serhii.omelchenko@nure.ua