

ОПТИМИЗАЦИЯ НИЗКОСКОРОСТНЫХ ЦИФРОВЫХ СИСТЕМ ПЕРЕДАЧИ РЕЧИ

В настоящее время проблема уплотнения и низкоскоростной передачи речи остается весьма актуальной. В системах уплотнения речи достигнута высокая степень сжатия размерности представления сигналов, что реализуется за счет ухудшения качества воспроизводимой речи. При современном уровне развития цифровой техники для уплотнения речевых сигналов наибольшее распространение получили методы, основанные на линейном предсказании сигналов. Суть этих методов состоит в том, что на этапе анализа оцениваются параметры модели источника возбуждения и параметры голосового тракта, представляемого в виде линейного фильтра. На этапе синтеза с помощью линейного фильтра формируется сигнал, близкий по слуховому восприятию к исходному речевому сигналу. В качестве параметров линейного фильтра наиболее широко используются коэффициенты линейного предсказания, коэффициенты отражения, нули и полюсы фильтра.

В данной работе предложена процедура оптимизации алгоритмов обработки речевых сигналов в телекоммуникационных системах по критерию максимальной разборчивости речи при заданной скорости цифрового потока в канале связи. При этом оптимизация систем передачи может выполняться в процессе функционирования систем, и тогда эта процедура будет относиться к классу адаптивных. В предложенной процедуре точность представления речевого сигнала на этапе анализа контролируется с помощью некоторой меры близости синтезированного и исходного сигналов. В качестве такой меры близости использован показатель разборчивости речи, обоснованный в работе [1]. Достоинством этого показателя является возможность его вычисления по реализациям речевых сигналов в отличие от традиционно используемых показателей разборчивости речи [2,3]. При этом, как показано в работе [1], при высокой разборчивости речи значения этих показателей оказываются тесно связанными.

Основные экспериментальные исследования посвящены задаче оптимизации алгоритмов кодирования сигналов для вокодерных систем, реализующих метод линейного предсказания. В качестве оцифровываемых параметров линейного фильтра использованы коэффициенты отражения, обладающие рядом достоинств: они просто оцениваются с помощью решетчатых фильтров, могут быть закодированы малым числом бит, их использование гарантирует устойчивость цифрового фильтра на этапе синтеза.

1. Постановка задачи оптимизации систем передачи речевых сообщений

Показатель формантной разборчивости речи широко применяется для анализа систем передачи речевых сообщений, так как он может быть непосредственно рассчитан на основе формантной теории разборчивости речи [1,2]. В данной работе для оптимизации систем передачи речевых сообщений использован модифицированный показатель формантной разборчивости [1]. Этот показатель описывает погрешность представления речевого сигнала в спектральной области с учетом особенностей восприятия речи человеком как некоторой совокупности формант. Он легко определяется по сегментам речевых сигналов с использованием хорошо разработанных алгоритмов линейного предсказания речи. С учетом указанных свойств этот показатель может быть использован для численной оптимизации систем обработки речевых сигналов.

Будем полагать, что структура системы передачи речевых сигналов известна. На ее вход поступает исходный речевой сигнал $x_0(t)$, а с выхода снимается переданный сигнал $x(t)$.

Сигнал $x(t)$ длительностью T_n на выходе системы передачи речи представим в виде совокупности $p = \lceil T_n / T_c \rceil$ следующих друг за другом коротких сегментов

$$x_r(t) = x(t + (r-1) \cdot T_c), \quad t \in [0, T_c); \quad r = \overline{1, p}. \quad (1)$$

Таким же образом представим и исходный речевой сигнал на входе системы

$$x_{or}(t) = x_0(t + (r-1) \cdot T_c), \quad t \in [0, T_c); \quad r = \overline{1, p}. \quad (2)$$

Состояние системы передачи определяется значением вектора параметров $\vec{\gamma}$. Множество Γ_0 допустимых значений этого параметра определяется конкретными особенностями прикладной задачи.

В качестве показателя разборчивости переданного речевого сигнала используем величину

$$\alpha(\vec{\gamma}) = \frac{1}{P} \sum_{r=1}^P \tilde{\alpha}_r(\vec{\gamma}), \quad (3)$$

в которой погрешности представления отдельных сегментов $\tilde{\alpha}_r(\vec{\gamma})$ определяются в соответствии с выражением

$$\tilde{\alpha}_r(\vec{\gamma}) = \frac{1}{J} \sum_{j=1}^J \frac{[S_r(\hat{f}_j, \vec{\gamma}) - S_{0r}(\hat{f}_j)]^2}{S_{0r}^2(\hat{f}_j)}, \quad (4)$$

где J – число формант речи, учитываемых на анализируемом сегменте; $\{\hat{f}_j, j = \overline{1, J}\}$ – оценки центральных частот формант речи на анализируемом сегменте, найденные по соответствующим сегментам исходного речевого сигнала с использованием одного из известных алгоритмов [4]; $S_{0r}(f), S_r(f, \vec{\gamma})$ – значения авторегрессионного спектра r -го сегмента для исходного и переданного сигнала соответственно. Вычисление значений авторегрессионных спектров в (4) осуществляется в соответствии с выражениями [4-6]:

$$S_{0r}(f) = \frac{\Delta t \cdot \sigma_{0r}^2}{\left| 1 + \sum_{m=1}^{M_0} a_{0m}^r \cdot \exp(-j2\pi f \cdot \Delta t \cdot m) \right|^2}, \quad S_r(f, \vec{\gamma}) = \frac{\Delta t \cdot \sigma_r^2(\vec{\gamma})}{\left| 1 + \sum_{m=1}^M a_m^r(\vec{\gamma}) \cdot \exp(-j2\pi f \cdot \Delta t \cdot m) \right|^2}, \quad (5)$$

где $\sigma_{0r}^2, \sigma_r^2(\vec{\gamma})$ – оценки дисперсии сигналов, возбуждающих голосовой тракт на r -ом сегменте; $\{a_{0m}^r, m = \overline{1, M_0}\}, \{a_m^r(\vec{\gamma}), m = \overline{1, M}\}$ – совокупности оценок коэффициентов авторегрессии для r -ого сегмента.

Найдем оптимальное значение вектора параметров системы передачи речи $\vec{\gamma}$ из условия минимума показателя $\alpha(\vec{\gamma})$ вида (3) по множеству допустимых значений Γ_0

$$\vec{\gamma}_0 = \arg \min_{\vec{\gamma} \in \Gamma_0} \{\alpha(\vec{\gamma})\}. \quad (6)$$

Описанный критерий оптимальности позволяет сформулировать и решить ряд конкретных задач по оптимизации систем передачи речевых сообщений. Определенный интерес представляют задачи оптимизации низкоскоростных систем передачи речи.

2. Особенности оптимизации низкоскоростных систем передачи

Рассмотрим системы передачи речи, в которых используются вокодеры линейного предсказания сигналов.

Будем полагать, что вокодерная система имеет известную структуру [4], изображенную на рис.1.

$$\sigma_r, \{K_{r,m}, m = \overline{1, M}\}, \nu_r$$

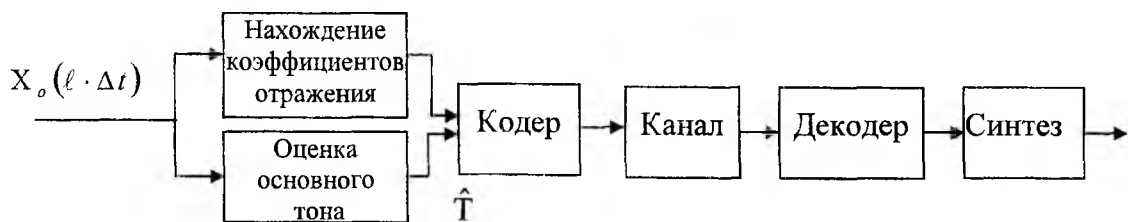


Рис. 1

На вход системы поступает последовательность отсчетов речевого сигнала $X_o(\ell \cdot \Delta t), \ell = \overline{1, L}$, где Δt - интервал дискретизации сигнала. Считается, что речь обрабатывается посегментно с длиной сегмента $T_c = \Delta t \cdot N$. Для этого отсчеты речевого сигнала $X_o(l \cdot \Delta t), \ell = 0, 1, \dots, L$ объединяются в сегменты, которые представлены векторами $\bar{X}_o^r = \{x_o(N \cdot r + \ell), \ell = \overline{0, N-1}, r = 0, 1, 2, \dots\}$, где Δt - интервал дискретизации; N - число отсчетов сигнала в сегменте.

По каждому сегменту \bar{X}_o^r вычисляются параметры: совокупность коэффициентов отражения $\{K_m^r, m = \overline{1, M}\}$; коэффициент усиления σ_r ; коэффициент вокализованности ν_r ; оценки периода основного тона \hat{T}_r . Квадрат коэффициента усиления σ_r^2 пропорционален мощности речевого сигнала, а коэффициент вокализованности характеризует вид источника возбуждения речевого тракта. Предполагается, что в оптимизируемой вокодерной системе для оценивания значений коэффициентов отражения используется решетчатый алгоритм [4, 6].

В кодере осуществляется в общем случае нелинейное квантование каждого из перечисленных $M+3$ параметров. Полученные числа представляются в двоичной системе исчисления и передаются по каналу связи. На приемном конце производится синтез сегментов речевых сигналов \bar{X}^r по общеизвестным алгоритмам [4].

Разборчивость речи на выходе вокодерной системы зависит от точности воспроизведения формант исходного сигнала после операций кодирования и декодирования. Численные значения показателя разборчивости синтезированной речи могут быть найдены в соответствии с выражением (3). При этом значения авторегрессионных спектров исходного $S_{0r}(\bar{f}_j)$ и синтезированного $S_r(\bar{f}_j, \bar{r})$ спектров вычисляются в соответствии с (5), для чего от совокупности коэффициентов отражения необходимо с помощью известных соотношений [4] перейти к совокупности оценок коэффициентов авторегрессии.

3. Оптимизация кодирования коэффициентов отражения

Здесь в рамках общей постановки задачи оптимизации систем передачи речевых сообщений рассматривается случай линейного кодирования и считается, что выполняются следующие условия:

1. Вокодерная система имеет известную структуру, изображенную на рис. 1.

2. Предполагается, что коэффициенты отражения для исходной речи имеют значения $K_m^r \in [K_{mn}, K_{ms}] m = \overline{1, M}, r = \overline{1, p}$, где K_{mn}, K_{ms} - соответственно нижняя и верхняя границы возможных значений коэффициентов ($-1 \leq K_{mn} < K_{ms} \leq 1$). Для конкретного абонента эти границы определяются на этапе предварительного анализа речевого сигнала. В кодере системы каждый из параметров (коэффициентов отражения) линейно квантуется на 2^{β_m} уровней и отображается в множество целых чисел $\{0, 1, \dots, 2^{\beta_m} - 1\} m = \overline{1, M}$, где β_m - число бит, используемых для представления соответствующего параметра.

Найдем оптимальное значение вектора параметров $\bar{\beta} = (\beta_1, \dots, \beta_M)$ из критерия

$$\alpha(\bar{\beta}_o) = \min_{\bar{\beta} \in B_o} \{\alpha(\bar{\beta})\}. \quad (7)$$

$$\alpha(\bar{\beta}) = \sum_{r=1}^n \alpha_r(\bar{\beta}), \quad (8)$$

где B_o - множество допустимых значений параметра, которые определяется равенством

$$\beta_1 + \beta_2 + \dots + \beta_M = b; \quad (9)$$

b - целое число, равное числу бит, выделенных на кодирование коэффициентов авторегрессии. Обычно $20 \leq b \leq 70$.

Отметим, что вид функции $\alpha_r(\bullet)$ задается выражением (4) с учетом указанных в разд.1 свойств вокодерных систем.

Особенности сформулированной задачи оптимизации состоят в следующем:

- она относится к задачам целочисленного программирования;
- ее непосредственное решение прямым перебором всех элементов множества допустимых решений B_∂ затруднительно, так как мощность данного множества равна числу возможных способов распределения b одинаковых объектов по M различным ячейкам [6]

$$|B_\partial| = C_{b+M-1}^{M-1} = \frac{(b+M-1)!}{(M-1)!b!}.$$

Например, при $b=25$ и $M=8$ имеем $|B_\partial| = C_{32}^7 = \frac{32!}{7!25!} \approx 3,4 \times 10^6$.

Отметим, что нахождение значения критерия (8) в каждой из $|B_\partial|$ точек требует значительного объема вычислений, выполнение которых на современных персональных компьютерах занимает несколько десятков секунд.

- Множество возможных значений параметра B_∂ может быть представлено в виде пересечения

$$B_\partial = G \cap Z_M,$$

где $Z^M \subset R^M$; $G = \{\bar{\beta} : (\beta_1, \dots, \beta_M) : \beta_1 + \dots + \beta_M = b\}$ - выпуклое множество $G \subset R^M$.

Выпуклость множества G следует из того, что если $\bar{\beta}' \in G$ и $\bar{\beta}'' \in G$, то для произвольного $\Theta \in [0,1]$ выполняется условие $\bar{\beta} = \Theta \bar{\beta}' + (1 - \Theta) \bar{\beta}'' \in G$, так как

$$\sum_{i=1}^M \beta_i = \sum_{i=1}^M [\Theta \beta_i' + (1 - \Theta) \beta_i''] = \Theta b + (1 - \Theta) b = b.$$

Учитывая особенности сформулированной задачи, воспользуемся для ее решения итерационным алгоритмом нелинейного программирования, основанном на схеме локальной оптимизации [8-10]. Его суть состоит в том, что вначале снимаются ограничения на параметры системы (9) и проводится процедура предварительной оптимизации. В результате этой процедуры последовательно уменьшается параметр скорости вокодера $\beta_1 + \dots + \beta_m$ до тех пор, пока не выполнится условие (9). Полученное значение вектора параметров используется затем в основном алгоритме оптимизации в качестве начального приближения.

Основной алгоритм оптимизации состоит в следующем.

Вначале выберем метрику на множестве B_∂ вида

$$r(\bar{\beta}', \bar{\beta}'') = \sum_{m=1}^M |\beta_m' - \beta_m''| \quad (10)$$

и определим окрестность радиуса ρ точки $\bar{\beta} \in B_\partial$ как

$$O_\rho(\bar{\beta}) = \{\bar{\beta}' : r(\bar{\beta}, \bar{\beta}') \leq \rho\}. \quad (11)$$

Положим радиус $\rho = 2$, так как $O_1(\bar{\beta}) \cap B_\partial = \emptyset$ для всех $\bar{\beta} \in B_\partial$.

Затем выберем некоторые начальные решения $\bar{\beta}_0 \in B_\partial$.

Считая, что на $(k-1)$ -м шаге было получено допустимое решение $\bar{\beta}_{k-1}$, находим на k -ом шаге ($k=1,2,\dots$) оптимальное в локальной задаче решение

$$\bar{\beta}_k = \arg \min \{\alpha(\bar{\beta}) : \bar{\beta} \in O_2(\bar{\beta}_{k-1}) \cap B_\partial\}. \quad (12)$$

Если $\alpha(\bar{\beta}_k) \leq \alpha(\bar{\beta}_{k-1})$, то необходимо перейти к следующему шагу, в случае же $\alpha(\beta_k) > \alpha(\beta_{k-1})$ значение β_{k-1} принимается в качестве локального минимума функции цели.

Достаточными условиями для глобальности найденного минимума являются выпуклость множества допустимых решений B_0 и выпуклость функции цели $\alpha(\bar{\beta})$ [9]. Выше было доказано, что первое из оговоренных условий в нашем случае выполняется. Поэтому сходимость алгоритма будет обеспечена, если функция цели удовлетворяет свойству выпуклости.

Доказать выполнение свойства выпуклости функции цели $\alpha(\bar{\beta})$, определяемой согласно (8), затруднительно из-за ее сложности. Однако поведение функции $\alpha(\bar{\beta})$, значение которой характеризует разборчивость речи, согласуются со свойством выпуклости. Эта функция является монотонно убывающей по каждой из координат и асимптотически стремится к нулю с ростом значений этих координат. Поэтому при достаточно высоком уровне разборчивости речи предположение о выпуклости функции $\alpha(\bar{\beta})$ можно считать физически обоснованным.

Для нахождения начального значения $\bar{\beta}_0$ в алгоритме (12) можно воспользоваться алгоритмом предварительной оптимизации, суть которого состоит в следующем.

Вначале выбирается начальное значение $\bar{\beta}_0$ с достаточно большим числом бит представления и показателем разборчивости $\alpha(\bar{\beta}_0) \approx 0$, что соответствует высокой разборчивости речи.

Затем выполняется рекуррентная процедура, на каждом шаге которой определяется значение $\bar{\beta}_k$

$$\bar{\beta}_k = \arg \min_{\bar{\beta}_k \in D} \alpha(\bar{\beta}), \quad (13)$$

где $D = O_1(\bar{\beta}_{k-1})$ – окрестность единичного радиуса точки $\bar{\beta}_{k-1}$, которая определяется согласно (11).

Процедура прекращается, когда

$$\sum_{m=1}^M \bar{\beta}_{k,m} = b,$$

т.е. когда число бит, используемых для кодирования, достигает заданного значения b .

Начальное значение $\bar{\beta}_0$ для алгоритма (12) полагается равным $\bar{\beta}_k$.

Описанный выше алгоритм оптимизации кодирования коэффициентов отражения для вокодерных систем реализован программно на языке "Pascal". Оптимизация проводилась отдельно для вокализованных и невокализованных сегментов речи, что обусловлено следующими соображениями. Вокализованные и невокализованные звуки речи имеют различную физическую природу и поэтому обладают различным распределением энергии по частоте. Для невокализованных звуков речи на представление коэффициентов отражения может быть выделено на 4-5 бит больше, чем для вокализованных, так как для них не требуется передача значений частоты основного тона.

Для вокализованных сегментов речи диктора мужского пола с использованием алгоритма (12) получено следующее оптимальное по введенному показателю распределение 25 бит между 8-ю коэффициентами отражения: $\beta_1 = 4$; $\beta_2 = 5$; $\beta_3 = 4$; $\beta_4 = 4$; $\beta_5 = 3$; $\beta_6 = 2$; $\beta_7 = 2$; $\beta_8 = 1$. При этом значение показателя разборчивости (8) составило $\alpha(\bar{\beta}) \approx 0,042$.

Для невокализованных сегментов речи того же диктора с использованием алгоритма (12) оптимальное распределение 30 бит между 8-ю коэффициентами отражения составило: $\beta_1 = 6$; $\beta_2 = 6$; $\beta_3 = 5$; $\beta_4 = 4$; $\beta_5 = 3$; $\beta_6 = 3$; $\beta_7 = 2$; $\beta_8 = 1$. При этом значение показателя разборчивости составило $\alpha(\beta) \approx 0,038$.

В приведенном примере задача оптимизации решалась при различном числе двоичных разрядов, требуемых для кодирования коэффициентов отражения на вокализованных и невокализованных сегментах речи. При этом для кодирования коэффициентов отражения на вокализованных сегментах отводилась на 5 двоичных разрядов меньше, чем на невокализованных сегментах. Эти разряды резервировались для представления частоты основного тона. Интересно отметить, что при указанном соотношении необходимых разрядов двоичного представления коэффициентов отражения были достигнуты примерно одинаковые значения показателя разборчивости речи для вокализованных и невокализованных сегментов.

4. Оптимизация нелинейного квантования коэффициентов отражения

В работе [4] авторы на основе экспериментальных исследований пришли к выводу, что следует применять нелинейное квантование коэффициентов отражения в силу их неоднородной спектральной чувствительности. Причем, наибольшая спектральная чувствительность будет в том случае, когда модуль коэффициента близок к единице. В указанных работах предлагается до линейного квантования выполнять нелинейное преобразование коэффициентов отражения вида

$$f(k_m) = \ln \frac{F - k_m}{F + k_m}, \quad m = \overline{1, M}, \quad (14)$$

где значение параметра F не превышает единицы.

Поэтому по аналогии с уже рассмотренной задачей была сформулирована задача оптимизации нелинейного преобразования коэффициентов отражения в классе преобразований вида (14) по критерию максимальной разборчивости речи с использованием показателя разборчивости (3).

Опуская детали решения указанной задачи, приведем лишь ее окончательный результат, суть которого состоит в следующем. Нелинейному преобразованию (14) следует подвергать только первый коэффициент отражения и только для вокализованных сегментов речи с параметром F близким к единице. Остальные коэффициенты отражения предпочтительно квантовать линейно. Такой вывод справедлив для случая, когда частота дискретизации сигнала в стандартном телефонном канале составляет 8 кГц и каждый сегмент сигнала представляется 8-ю коэффициентами отражения.

Заключение

В настоящей статье сформулирована общая постановка задачи оптимизации алгоритмов обработки речевых сигналов для цифровых систем передачи по критерию максимальной разборчивости речи. Рассмотрены особенности задач оптимизации низкоскоростных систем передачи речи. Решены типовые задачи оптимизации алгоритмов обработки речевых сигналов для вокодеров с линейным предсказанием, использующих в качестве информативных параметров коэффициенты отражения. Разработан алгоритм оптимизации линейного кодирования коэффициентов отражения, который относится к классу алгоритмов целочисленного программирования, исследованы условия его сходимости. Приведены результаты решения задач оптимизации низкоскоростных систем передачи с вокодерами линейного предсказания.

Список литературы: 1. Пресняков И.Н., Омельченко А.В., Пресняков А.И. Оценка качества речи в цифровых системах передачи. Радиотехника. 2001. Вып. 120. С. 170-174. 2. Покровский Н. В. Расчет и измерение разборчивости речи. М.: Связьиздат, 1962. 390 с. 3. Вемян Г.В. Передача речи по сетям электросвязи. М.: радио и связь, 1985. 272с. 4. Дж. Д. Маркел, А. Х. Грей. Линейное предсказание речи. М.: Связь, 1980. 308с. 5. Рабинер Л. Р., Шафер Р. В. Цифровая обработка речевых сигналов /Под ред. М. В. Назарова и Ю. Н. Прохорова. М.: Радио и связь, 1981. 496 с. 6. Марпл. - мл. С. Л. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 584 с. 7. Глушкин Л. М., Шварц В. Я, Шор Л. А. Задачи и алгоритмы комбинаторики и теории графов. Донецк.: ДПИ, 1982. 110 с. 8. Ковалев М. М. Дискретная оптимизация (целочисленное программирование). М.: Изд-во БГУ, 1977. 190 с. 9. Сергиенко И. В., Лебедева Т. Т., Роцин В. А. Приближенные методы решения дискретных задач оптимизации. Киев.: Наук. думка, 1980. 274 с. 10. Финкельштейн Ю. Ю. Приближенные методы и прикладные задачи дискретного программирования. М.: Наука, 1976. 264 с.