

УДК 534.78+621.391

# НОВЫЙ ПОДХОД К ВЫЧИСЛЕНИЮ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ЧАСТОТ РЕЧЕВЫХ СИГНАЛОВ, ОСНОВАННЫЙ НА СВОЙСТВЕ МЕЖФРЕЙМОВОЙ УПОРЯДОЧЕННОСТИ

В. Ю. СЕМЕНОВ

Государственное научно-производственное предприятие “Дельта”, Киев

Получено 01.11.2004

Рассмотрена актуальная задача поиска линейных спектральных частот (ЛСЧ), являющихся альтернативной формой представления полюсной модели голосового тракта. Проведено исследование взаимного расположения ЛСЧ на смежных квазистационарных временных интервалах (фреймах). На основании этого предложен новый подход к вычислению ЛСЧ, состоящий из двух этапов – локализации ЛСЧ и их точного вычисления. Показано, что этап локализации эффективно сводится к проверке свойства межфреймовой упорядоченности ЛСЧ. Продемонстрировано, как для ускорения стадии точного вычисления ЛСЧ могут быть использованы их значения, найденные на предыдущем фрейме. В результате тестирования различных речевых сигналах продемонстрировано сокращение количества операций до 2.5 раз по сравнению с версией метода, не использующей априорную информацию о взаимном расположении ЛСЧ на смежных фреймах, и до 3.4 раз по сравнению с широко используемым методом Кабала. Кроме того, показано, что максимальные (пиковые) вычислительные затраты предложенного метода меньше не только минимальных затрат метода Кабала, но и минимальных затрат ускоренной комбинации метода Кабала с методом Ньютона. Это свидетельствует о предпочтительности использования разработанного метода в системах реального времени.

Розглянуто актуальну проблему пошуку лінійних спектральних частот (ЛСЧ), які є еквівалентною формою представлення полюсної моделі голосового тракту. Виконано дослідження взаємного розташування ЛСЧ на суміжних квазістаціонарних часових інтервалах (фреймах). На основі цього запропоновано новий підхід до обчислення ЛСЧ, який складається з двох етапів – локалізації ЛСЧ та їх точного обчислення. Показано, що етап локалізації ефективно зводиться до перевірки властивості міжфреймової упорядкованості ЛСЧ. Продемонстровано, як для прискорення стадії точного обчислення ЛСЧ можуть бути використані їхні значення, знайдені на попередньому фреймі. У результаті тестування на різноманітних мовних сигналах продемонстровано зменшення кількості операцій до 2.5 разів у порівнянні з версією методу, який не використовує априорну інформацію про взаємне розташування ЛСЧ на суміжних фреймах, та до 3.4 разів у порівнянні з широко використовуваним методом Кабала. Крім того, показано, що максимальні (пікові) обчислювальні витрати запропонованого методу менші не тільки за мінімальні витрати методу Кабала, але й за мінімальні витрати прискореної комбінації методу Кабала з методом Ньютона. Це свідчить про перевагу застосування розробленого методу в системах реального часу.

The paper is dedicated to an important problem of calculation of line spectral frequencies (LSF), being the alternative form of representation of all-pole vocal tract model. Investigation of LSF relative position on the adjacent quasi-stationary time intervals (frames) is performed. On this basis a new two-stage approach to LSF calculation is proposed. This approach consists of two parts: LSF localization and their refinement. It was shown that the localization stage is effectively reduced to verification of interframe ordering property. It is also demonstrated how LSF values obtained in the previous frame can be used to accelerate the LSF refinement stage. As a result of testing on various speech signals, the 2.5 times reduction of the number of operations is shown in comparison with the similar algorithm without allowance for a priori information about relative LSF position on the adjacent frames. It was also shown that computational expenses are reduced 3.4 times in comparison with widely used Kabal's method. Moreover, it is demonstrated that the maximum (peak) computational expenses of the proposed method are lower than minimal ones both for Kabal's method and the accelerated combination of Kabal's method with Newton's method. These results show the advantage of applying the proposed method in real-time systems.

## ВВЕДЕНИЕ

Современные отрасли речевой акустики, такие как распознавание речи, кодирование (сжатие) речи, верификация и идентификация диктора, речевой синтез, коррекция речевых сигналов, требуют применения простых и в то же время эффективных параметрических моделей полезного сигнала, основанных на сведениях из физики и физиологии образования речи, а также на особенностях слухового восприятия. Как правило, процесс образования голосовых звуков рассматривается как резуль-

тат прохождения возбуждающего процесса, моделирующего звуковое давление на выходе голосовой щели, через фильтр с изменяющимися во времени параметрами. Последний и представляет собой модель голосового тракта. Большинство современных методов цифровой обработки речевых сигналов основаны на использовании авторегрессионной (АР) модели образования речи [1, 2], в которой голосовой тракт моделируется посредством полюсного фильтра.

Упрощенная схема основных преобразований, которые претерпевает речевой сигнал при ра-

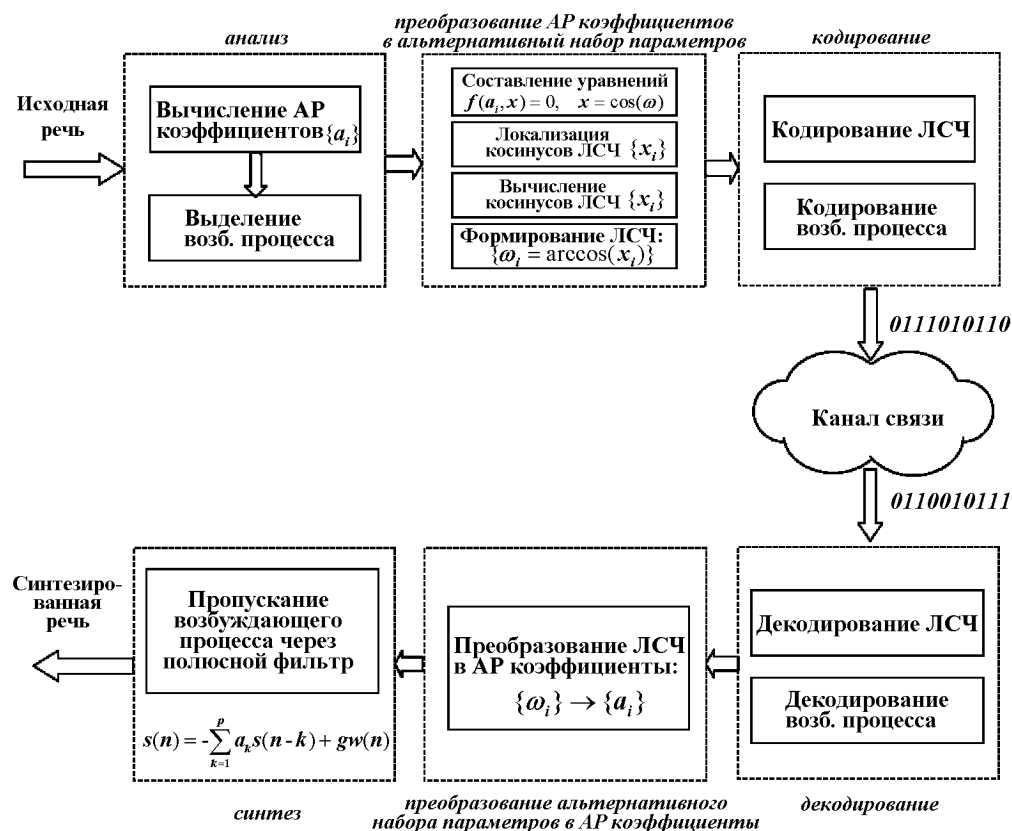


Рис. 1. Структура алгоритма преобразования речевых сигналов в типовых современных устройствах цифровой телефонии

боте типовых современных устройств голосовой связи (систем мобильной телефонии [3], IP-телефонии [4–6], вокодеров [7]), приведена на рис. 1.

Несмотря на то, что фильтр, моделирующий работу голосового тракта, однозначно определяется набором АР коэффициентов, в практических приложениях непосредственно АР коэффициенты не используются. В первую очередь, это объясняется их чрезвычайно высокой спектральной чувствительностью и отсутствием четких динамических диапазонов изменения. Таким образом, возникает потребность во введении промежуточного этапа, связанного с преобразованием АР коэффициентов в некоторый альтернативный набор параметров, для которых можно применить эффективные процедуры кодирования (см. рис. 1). По аналогичным причинам необходимо преобразовывать АР коэффициенты в альтернативные наборы параметров в задачах распознавания речи, идентификации диктора и других приложениях (см. обзор [10]).

На протяжении нескольких десятилетий в ка-

честве эквивалентных представлений коэффициентов АР модели использовались коэффициенты отражения и другие величины, связанные с моделью голосового тракта в виде последовательности соосных труб [1, 8]. Однако наиболее популярным способом альтернативного представления АР параметров стало их преобразование в набор линейных спектральных частот. Впервые введенные в 1975 году Итакура [12], ЛСЧ<sup>1</sup> постепенно получили широчайшее внедрение в алгоритмах кодирования [14–17], распознавания речи [18], идентификации диктора [19], методах коррекции искаженных речевых сигналов [9, 20]. Кроме того, они стали основой для введения мер искажения речевых сигналов [17, 18, 21].

Значительное количество современных устройств обработки речевых сигналов предназначено для работы в режиме реального времени. Однако нахождение ЛСЧ связано с поиском

<sup>1</sup>В англоязычной литературе используются эквивалентные термины “line spectral frequencies” (LSF) и “line spectrum pairs” (LSP).

корней уравнений. Это является нежелательным элементом для большинства вычислительных средств (особенно сигнальных процессоров с фиксированной точкой), поскольку может приводить к непредсказуемым задержкам и накоплению погрешностей округления. Проблемы вычисления ЛСЧ привлекали внимание многочисленных исследователей [13, 22–24] (подробнее см. также обзорную часть [10]). В связи с этим, в работе [10] были предложены новые методы поиска ЛСЧ, основанные на разработанном алгоритме решения трансцендентных уравнений. Хотя предложенный подход имеет ряд значительных преимуществ перед существующими аналогами, в нем не использованы физические особенности ЛСЧ, в которых должен заключаться резерв повышения эффективности метода. Заметим, что игнорирование априорной информации о распределении ЛСЧ является слабой стороной большинства традиционных алгоритмов их поиска.

Исходя из этого, цель данной работы состояла в построении экономичного метода вычисления ЛСЧ, максимально просто и эффективно использующего особенности их временного распределения. После изложения в первом разделе физической сути ЛСЧ, во втором разделе проводится исследование взаимного расположения ЛСЧ на смежных квазистационарных временных интервалах. В третьем разделе на основании сделанных выводов формулируется новый метод вычисления ЛСЧ, основанный на свойстве межфреймовой упорядоченности. В четвертом разделе, посвященном экспериментальным исследованиям, на основании различных критериев проводится сравнение эффективностей предложенного подхода, метода, изложенного в работе [10], традиционного метода Кабала [13] и его оптимизированной модификации. В конце статьи кратко сформулированы выводы и приведен список использованной литературы.

## 1. СВЯЗЬ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ЧАСТОТ С ПОЛЮСНОЙ МОДЕЛЬЮ ГОЛОСОВОГО ТРАКТА

Авторегрессионная (полюсная) модель голосового тракта рассматривает речевой сигнал как результат прохождения возбуждающего процесса через полюсную фильтр вида

$$H_s(z) = \frac{g}{A(z)} = \frac{g}{1 + \sum_{k=1}^p a_k z^{-k}}, \quad (1)$$

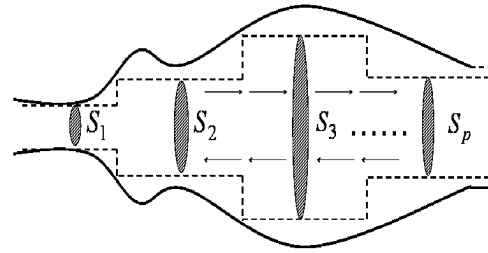


Рис. 2. Представление голосового тракта в виде последовательности труб

где  $a_k$ ,  $k=1, 2, \dots, p$  – коэффициенты, определяющие форму голосового тракта при произнесении звуков;  $g$  – коэффициент усиления, характеризующий уровень звука. При этом считается, что параметры полюсной модели голосового тракта являются неизменными на интервалах порядка 20 мс (свойство квазистационарности, обусловленное анатомическими ограничениями на динамику движения артикуляционных органов)<sup>2</sup>.

В классических работах [1, 2] показано, что модель (1) эквивалентна представлению голосового тракта в виде последовательного соединения труб одинаковой длины и различных диаметров (рис. 2). Записывая уравнения распространения звука в каждой из труб и учитывая граничные условия на их границах, можно показать, что передаточная функция такой системы имеет вид (1) (символ  $z$ -преобразования связан с символом преобразования Лапласа  $s$  соотношением  $z = \exp(2ls/c)$ , где  $l$  – длина трубы-секции;  $c$  – скорость звука). При этом коэффициенты  $\{a_k\}$  многочлена  $A(z)$ , стоящего в знаменателе передаточной функции (1), могут быть рассчитаны по рекуррентным соотношениям

$$\begin{aligned} A_0(z) &= 1, & A(z) &= A_p(z), \\ A_i(z) &= A_{i-1}(z) + k_i z^{-i} A_{i-1}(z^{-1}), & (2) \\ & & i &= 1, \dots, p. \end{aligned}$$

Здесь параметры  $k_i$ , получившие название коэффициентов отражения<sup>3</sup>, выражаются простым образом через площади поперечного сечения труб  $S_i$ :

$$k_i = \frac{S_{i+1} - S_i}{S_{i+1} + S_i}, \quad i = 1, \dots, p.$$

<sup>2</sup>При дальнейшем изложении временные интервалы, длина которых равна периоду квазистационарности речевого сигнала, будем называть фреймами.

<sup>3</sup>Физический смысл коэффициентов отражения состоит в определении величины волны, отраженной на границе двух труб.

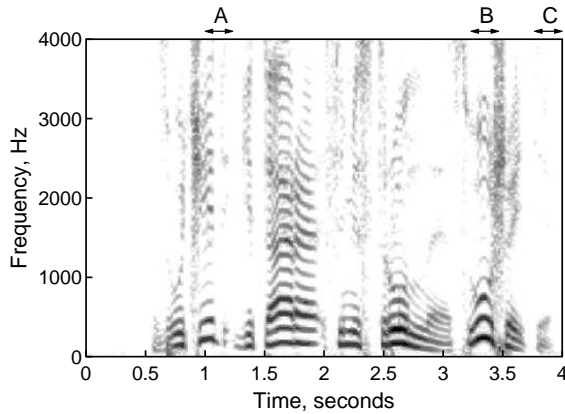


Рис. 3. Спектрограмма тестового речевого сигнала “Звучит гитара семиструнная. Слушайте!”

Исходя из соотношений (2), можно сформировать два искусственных полинома  $(p+1)$ -го порядка, положив  $(p+1)$ -ый коэффициент отражения  $k_i$  равным 1 и  $-1$  соответственно:

$$\begin{aligned} A_{p+1}^{(1)}(z) &= A_p(z) + z^{-p-1}A_p(z^{-1}), \\ A_{p+1}^{(2)}(z) &= A_p(z) - z^{-p-1}A_p(z^{-1}). \end{aligned} \quad (3)$$

Эти два случая соответствуют полному закрытию или полному открытию голосовой щели [12]. Было показано [12], что корни полиномов (3) лежат на единичной окружности. Их аргументы получили название линейных спектральных частот. Замечательной особенностью ЛСЧ является свойство внутрифреймовой упорядоченности:

$$0 < \omega_1 < \omega_2 < \omega_3 < \dots < \omega_{p-1} < \omega_p < \pi \quad (4)$$

(нечетные ЛСЧ соответствуют первому из полиномов (3), а четные – второму).

Помимо свойства внутрифреймовой упорядоченности (4) и ограниченного диапазона изменения, другими важными свойствами ЛСЧ оказываются их тесная связь с формантными частотами, локализованная спектральная чувствительность, а также межфреймовая коррелированность, заключающаяся в статистической зависимости ЛСЧ для смежных временных интервалов (см. обзор [17]). Наличие такой зависимости позволяет сформулировать вопрос: могут ли определенные на некотором временном интервале значения ЛСЧ быть эффективно использованы для вычисления ЛСЧ следующего временного интервала?

## 2. ИССЛЕДОВАНИЕ ВЗАИМНОГО РАСПОЛОЖЕНИЯ ЛСЧ НА СМЕЖНЫХ ФРЕЙМАХ

Исходя из связи ЛСЧ с акустической моделью голосового тракта и, в особенности, учитывая их тесную взаимосвязь с его резонансами [17], можно сделать предположение, что ЛСЧ, соответствующие смежным временным интервалам, должны относительно мало отличаться друг от друга. Для того, чтобы исследовать справедливость данного предположения, рассмотрим речевой сигнал продолжительностью 4.62 с, произнесенный диктором-мужчиной и оцифрованный с частотой дискретизации  $f_s = 8000$  Гц. Его спектрограмма отображена на рис. 3.

Для иллюстрации взаимного расположения ЛСЧ на смежных временных интервалах рассмотрим три фрагмента сигнала, отмеченные на рис. 3 буквами А, В и С. Временные реализации этих фрагментов и соответствующие им графики ЛСЧ приведены на рис. 4, 5 и 6 соответственно. Вычисление ЛСЧ проводилось на последовательных интервалах длиной 20 мс при порядке АР модели голосового тракта  $p = 10$ . Для облегчения интерпретации получаемых распределений ЛСЧ применялась их нормировка:  $f_k = \omega_k f_s / (2\pi)$ ,  $k = 1, \dots, p$ .

На рис. 4 представлена ситуация, соответствующая завершению звука “и” и началу звука “т” в слове “звучит”. Анализ рисунка показывает, что, несмотря на изменение всех ЛСЧ в достаточно широких пределах, практически во всех случаях ЛСЧ с номером  $i$  находится между частотами предыдущего фрейма с номерами  $i-1$  и  $i+1$ . Единственное исключение составляет третья ЛСЧ, принадлежащая седьмому из рассмотренных интервалов (выделена кружочком), которая немного (на 4 Гц) превосходит четвертую ЛСЧ предыдущего временного фрейма. Как видно из представленной на рис. 4 временной реализации данного речевого фрагмента, эта ситуация соответствует началу нового звука (согласной “т”).

На рис. 5 показано распределение ЛСЧ на протяжении непрерывного согласного звука “л” (в слове “слушайте”) с характерным сближением ЛСЧ в области основного тона. Такое поведение объясняется характерным свойством ЛСЧ, заключающемся в окаймлении наиболее ярко выраженных спектральных пиков. Как видно из графика, на всех проанализированных интервалах данного речевого фрагмента ЛСЧ  $\omega_i$  находится между частотами предыдущего фрейма с номерами  $i-1$  и  $i+1$ , т. е.

выполняется неравенство

$$\omega_{i-1}^{(n-1)} < \omega_i^{(n)} < \omega_{i+1}^{(n-1)}, \quad i = 2, \dots, p-1 \quad (5)$$

(верхние индексы обозначают номера фреймов).

Рис. 6 иллюстрирует завершение финального гласного звука “е” и последующей паузы, заполненной фоновым шумом. Как следует из рисунка, на участке, соответствующем затуханию гласного звука, ЛСЧ изменяются в достаточно широких пределах. Тем не менее, ЛСЧ с номером  $i$  постоянно находится между частотами предыдущего фрейма с номерами  $i-1$  и  $i+1$ . На участке, соответствующем присутствию фонового шума, ЛСЧ имеют равномерное распределение, практически инвариантное во времени и, как следствие, также удовлетворяют неравенству (5). Следует отметить, что паузы обычно составляют не менее 40÷50 % продолжительности речевых сигналов. Это обстоятельство дополнительно усиливает предположение о том, что соотношение (5) справедливо в большинстве встречающихся на практике ситуаций.

Для подтверждения этой гипотезы выполнение неравенства (5) было проверено на базе тестовых речевых записей девяти дикторов общей продолжительностью около 8 мин. Предварительно из данных сигналов (оцифрованных с частотой дискретизации  $f_s = 8000$  Гц) удалили участки, содержащие паузы в начале и конце фраз. Это позволило максимально объективно оценить нижнюю границу вероятности выполнения неравенства (5). В ходе тестирования были рассмотрены наиболее распространенные порядки АР моделей: от 8 до 20. В табл. 1 представлена процентная доля случаев, в которых неравенство (5) выполнялось для всех ЛСЧ текущего фрейма ( $n_0$ ), а также доли случаев, в которых данное неравенство нарушалось в одном ( $n_1$ ), двух ( $n_2$ ) и трех ( $n_3$ ) случаях соответственно.

Как видно из таблицы, соотношение (5) выполняется в подавляющем большинстве ситуаций. Соответствующее процентное отношение составляет от 88.85 % при  $p=20$  до 96.78 % при  $p=8$ . Как следствие, в таких случаях аналогичным свойством обладают и косинусы ЛСЧ (взятые в обратном порядке):

$$x_{i-1}^{(n-1)} < x_i^{(n)} < x_{i+1}^{(n-1)}, \quad i = 2, \dots, p-1. \quad (6)$$

Таким образом, в преобладающем количестве случаев аргументы корней каждого из уравнений (3) лежат между вычисленными на предыдущем фрейме аргументами корней соответствующего уравнения-пары. Это позволяет утверждать,

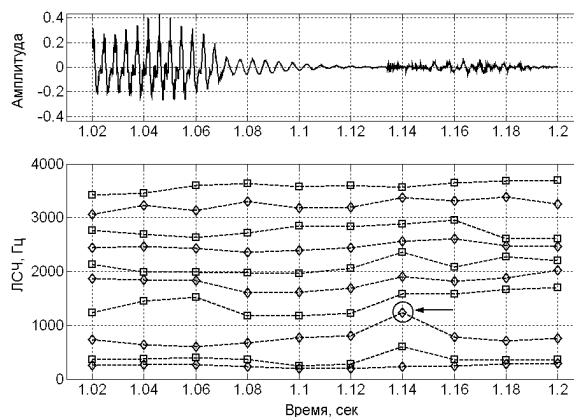


Рис. 4. Распределение ЛСЧ на границе двух звуков

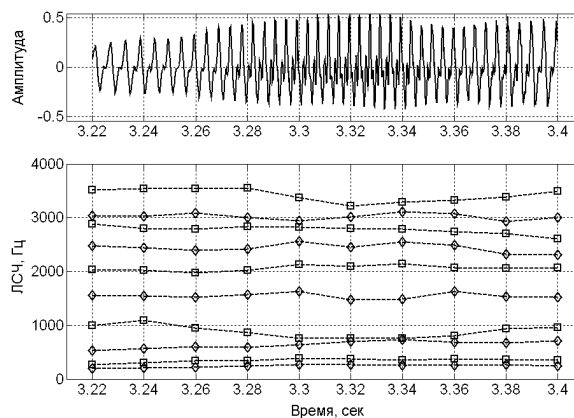


Рис. 5. Распределение ЛСЧ на протяжении звонкого звука

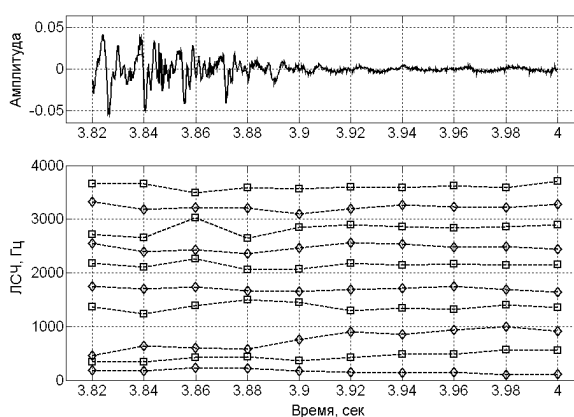


Рис. 6. Распределение ЛСЧ, соответствующее завершению гласного звука и последующей за ним паузы

Табл. 1. Процентные доли случаев выполнения ( $n_0$ ) и невыполнения ( $n_1, n_2, n_3$ ) условия межфреймовой упорядоченности (5) при различных порядках моделей

	$p=8$	$p=10$	$p=12$	$p=14$	$p=16$	$p=18$	$p=20$
$n_0$	96.78	95.62	94.36	93.17	91.95	90.14	88.85
$n_1$	2.81	3.67	4.61	5.44	6.37	7.62	8.19
$n_2$	0.33	0.54	0.78	0.95	1.12	1.47	1.97
$n_3$	0.07	0.15	0.19	0.27	0.35	0.45	0.55

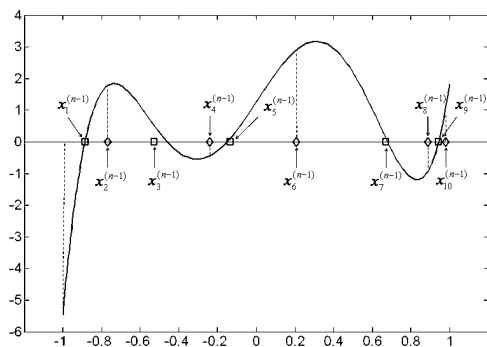


Рис. 7. Пример локализации косинусов ЛСЧ с помощью значений, вычисленных на предыдущем фрейме

что задача локализации ЛСЧ сводится, главным образом, к проверке условия их межфреймовой упорядоченности (5) или (6).

### 3. ПРЕДЛАГАЕМЫЙ АЛГОРИТМ ВЫЧИСЛЕНИЯ ЛИНЕЙНЫХ СПЕКТРАЛЬНЫХ ЧАСТОТ

Исходя из проведенного исследования взаимного расположения ЛСЧ на смежных фреймах, предлагается следующий алгоритм их вычисления (соответствующая последовательность этапов приведена в блок-схеме рис. 1). При этом предполагается, что перед вычислением ЛСЧ на временном интервале с номером  $n$  известны значения косинусов ЛСЧ  $x_1^{(n-1)}, x_2^{(n-1)}, \dots, x_{p-1}^{(n-1)}, x_p^{(n-1)}$ , вычисленные на предыдущем фрейме.

#### 3.1. Сведение уравнений (3) к полиномиальному виду

Путем замены  $x = \cos(\omega)$  соотношения (3) сводятся к полиномиальным уравнениям степени  $M = p/2$  (подробные выкладки приведены в статье [10]):

$$\sum_{k=0}^M c_{m,k} x^{M-k} = 0, \quad m = 1, 2. \quad (7)$$

#### 3.2. Локализация косинусов ЛСЧ

Проверка выполнения условия (6) может быть осуществлена путем сравнения знаков функций (7) в точках

$$\{-1, x_2^{(n-1)}, x_4^{(n-1)}, \dots, x_{p-2}^{(n-1)}, x_p^{(n-1)}, 1\}$$

при вычислении нечетных корней

$$\{x_1^{(n)}, x_3^{(n)}, \dots, x_{p-3}^{(n)}, x_{p-1}^{(n)}\}$$

или на сетке

$$\{-1, x_1^{(n-1)}, x_3^{(n-1)}, \dots, x_{p-3}^{(n-1)}, x_{p-1}^{(n-1)}, 1\}$$

при вычислении четных корней. Иллюстрацией к этому этапу является рис. 7, на котором приведен типичный график первого из полиномов (7) и показаны косинусы ЛСЧ  $\{x_i^{(n-1)}\}$ , вычисленные на предыдущем фрейме. Как видно из рисунка, рассматриваемая функция последовательно меняет знак в точках

$$\{-1, x_2^{(n-1)}, x_4^{(n-1)}, \dots, x_{p-2}^{(n-1)}, x_p^{(n-1)}, 1\}.$$

Таким образом, необходимое для локализации корней каждого из полиномов (7) количество вызовов полиномиальных функций составляет всего  $p/2+1$  (что контрастирует с количеством вызовов для метода Кабала [13], превышающим  $p^2/2$ ).

В том случае, когда для некоторого номера  $i$  условие (6) не выполняется в том смысле, что функция (7) имеет одинаковые знаки на краях отрезка  $[x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}]$ , проверка наличия корня осуществляется с помощью алгоритма, предложенного в [10]. Указанный алгоритм проверяет наличие корней на данном отрезке с помощью хорошо разработанного инструментария анализа функции и ее производной, при необходимости проводя рекурсивное деление исходного отрезка.

#### 3.3. Уточнение значений ЛСЧ

После завершения этапа локализации на каждом из выделенных отрезков точные значения

Табл. 2. Сравнение средних вычислительных затрат (MFlops), требуемых методом [10] и методом, предложенным в данной работе

Алгоритм	$p=8$	$p=10$	$p=12$	$p=14$	$p=16$	$p=18$	$p=20$
Метод [10]	0.046	0.075	0.124	0.172	0.236	0.314	0.381
Предложенный метод	0.029	0.044	0.061	0.082	0.104	0.129	0.158

косинусов ЛСЧ определяются с помощью одной из стандартных итерационных процедур, например, метода Ньютона [25, 26]. При этом для нахождения корня  $x_k^{(n)}$  предлагается в качестве начального приближения использовать вычисленное на предыдущем фрейме значение  $x_k^{(n-1)}$ . Эта ситуация наглядно иллюстрируется с помощью рис. 7. Действительно, нечетные косинусы ЛСЧ, вычисленные на предыдущем фрейме, не только являются хорошим приближением для нечетных корней текущего фрейма, но в ряде случаев практически совпадают с ними. На завершающем этапе ЛСЧ  $\omega_k$  формируются из полученных корней путем преобразования  $\omega = \arccos(x)$ .

#### 4. ЭКСПЕРИМЕНТАЛЬНЫЕ РЕЗУЛЬТАТЫ

Как известно, различные приложения цифровой обработки речевых сигналов требуют использования разных порядков АР модели  $p$ . Поэтому эффективность предложенного подхода проверена при наиболее часто употребляемых значениях  $p$ : 8, 10, ..., 20. Для возможности качественного и количественного оценивания изменений первоначально использовалась та же база тестовых сигналов, что и в работе [10]. Она состояла из пятиминутных записей шести дикторов (четырех мужчин и двух женщин), дискретизированных с частотой  $f_s = 8000$  Гц. Для каждого из исследуемых порядков АР модели речевые сигналы разбивались на фреймы длиной 20 мс (160 дискретных отсчетов). На каждом из этих интервалов с помощью автокорреляционного метода линейного предсказания [1, 2] были предварительно определены АР коэффициенты. Затем для каждого из этих наборов подсчитывались ЛСЧ.

В табл. 2 представлены средние количества выполняемых элементарных операций (в миллионах элементарных операций в секунду, MFlops), соответствующие методу, описанному в работе [10], и методу, предложенному в разделе 3. Для объективности сравнения на этапе уточнения значений косинусов ЛСЧ использовался традиционный метод Ньютона с таким же способом инициализации, как и в [10]. В качестве критерия оста-

новки алгоритма выбрано условие  $|f(x)| < 10^{-6}$ . Из таблицы видно, что предлагаемый метод обеспечивает экономию вычислительных затрат по сравнению с методом [10] от 1.6 (при  $p=8$ ) до 2.4 (при  $p=10$ ) раз. Столь существенное сокращение вычислительных затрат обусловлено введением более совершенной процедуры локализации ЛСЧ.

Затраты метода, предложенного в [10], в значительной степени определялись числом рекурсивных разбиений исходного диапазона поиска ЛСЧ. Следует заметить, что использование рекурсивных процедур (даже если оно и не вызывает существенных временных задержек) приводит к загроможденности стека и является нежелательным для устройств, работающих в режиме реального времени. В связи с этим замечательно то, что предложенный в данной статье метод требует проводить разбиения исходных отрезков поиска корней лишь в том случае, когда не выполняется условие (6), что, как уже обсуждалось выше, происходит достаточно редко. При наиболее часто используемом порядке АР модели  $p=10$  среднее количество разбиений, приходящихся на один фрейм, составило всего 0.076. Это резко контрастирует с соответствующим значением 9.97 (т. е. большим в 131 раз) для метода [10]. Таким образом, рекурсивные разбиения теперь приходится производить в среднем всего лишь один раз за 13 фреймов (порядка 0.65 секунды)! Это гарантирует отсутствие задержек, важное для реализации алгоритма в системах реального времени.

Кроме того, поскольку тестовые речевые сигналы отличались разнообразием сменяющих друг друга дикторов и практически полным отсутствием пауз, приводимые данные для вычислительных характеристик разработанного метода следует воспринимать, как соответствующие верхние границы. В процессе реальной работы устройств цифровой обработки речи значения вычислительных затрат и рекурсивных разбиений должны быть еще более низкими<sup>4</sup>.

Теперь сопоставим эффективность предложен-

<sup>4</sup>Во многих приложениях цифровой обработки речевых сигналов ЛСЧ вычисляются на перекрывающихся временных интервалах, что делает применение предложенного подхода еще более выгодным.

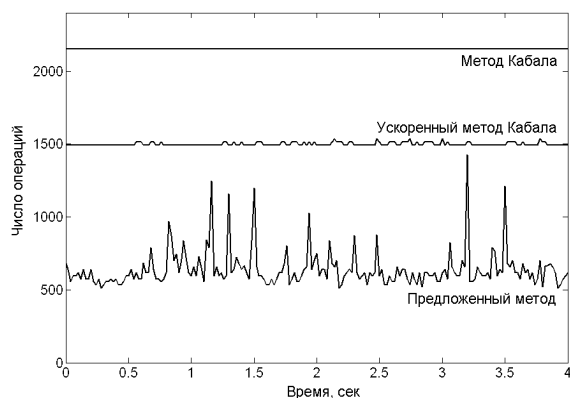


Рис. 8. Пример распределения вычислительных затрат методов поиска ЛСЧ в зависимости от времени. Пики в распределении операций предложенного метода соответствуют резким переходам от одного звука к другому

ного метода с распространенным сеточным методом Кабала [13] при определении ЛСЧ для наиболее часто используемого порядка АР модели  $p=10$ . Основные особенности использования метода Кабала подробно проанализированы в [10]. Поскольку в алгоритме Кабала используется уточнение косинусов ЛСЧ с помощью метода половинного деления (для фиксированности количества операций), для объективности сравнения была рассмотрена также его “ускоренная” модификация, соответствующая уточнению корней методом Ньютона. При этом выяснено, что наилучший результат этот комбинированный подход обеспечивает при инициализации метода Ньютона с помощью одной итерации метода ложного положения [25, 26].

В табл. 3 представлены средние, минимальные и максимальные количества операций, необходимые для выполнения на одном фрейме метода Кабала, его “ускоренной” модификации, метода [10] и метода, предлагаемого в данной работе. В качестве признака завершения поиска корней взято требуемое методом Кабала условие  $|x_k - x_{k-1}| < \epsilon$ , где  $x_k, x_{k-1}$  – приближенные значения корня, полученные на двух последовательных итерациях. Значение порога  $\epsilon$  взято равным  $10^{-3}$ . В данной серии экспериментов использовался речевой массив, описанный в разделе 2 и содержащий в качестве подмножества базу тестовых сигналов, применявшихся на предыдущем этапе исследования.

Из табл. 3 следует, что предложенный подход к поиску ЛСЧ обеспечивает сокращение средних вычислительных затрат в 3.42 раза по сравнению с методом Кабала. Также имеется достаточно су-

щественный (в 2.38 раза) выигрыш по сравнению с комбинацией метода Кабала и метода Ньютона. Обеспечиваемый выигрыш в 2.08 раза по сравнению с методом [10] согласуется с данными, приведенными в табл. 2.

Один из важных результатов работы [10] состоял в том, что помимо выигрыша по сравнению с методом Кабала в среднем количестве операций, максимальные (пиковые) количества элементарных операций, зафиксированные в процессе реальной работы, были ниже соответствующих затрат метода Кабала. Поэтому значительный интерес представляет исследование пиковых вычислительных затрат, свойственных предлагаемому методу поиска ЛСЧ. Как следует из табл. 3, он обладает еще более существенным преимуществом: соответствующее ему *максимальное* количество элементарных операций (1428) является меньшим не только затрат “прямого” метода Кабала (2150), но и *минимальных* затрат ускоренной комбинации метода Кабала с методом Ньютона (1491 элементарная операция). Данное обстоятельство является дополнительным свидетельством предпочтительности предложенного метода для применения в системах реального времени.

Дополнительный пример к сравнительной характеристике различных методов поиска ЛСЧ дан на рис. 8. Здесь изображены временные зависимости требуемого количества операций для поиска ЛСЧ тестового речевого сигнала, спектрограмма которого была представлена на рис. 3. Из графика наглядно следует преимущество предложенного метода в контексте выполняемого количества операций. При этом минимальные вычислительные затраты имеют место в паузах, заполненных стационарным фоновым шумом, а характерные пики в распределении операций соответствуют резким переходам от одного звука к другому.

В завершение следует заметить, что еще более значительное сокращение вычислительных затрат может быть достигнуто при внедрении в структуру предложенного метода следующих усовершенствований:

- а) применение правила Декарта [24] для ускорения локализации положительных и отрицательных корней многочленов (7);
- б) применение более быстро сходящихся методов поиска корней полиномов (7) [22].

## ЗАКЛЮЧЕНИЕ

Основной задачей данной работы было построение простого и эффективного метода вычисле-



Табл. 3. Сравнение средних и экстремальных значений вычислительных затрат для различных методов поиска ЛСЧ

Алгоритм	Среднее кол-во операций	Минимальное кол-во операций	Максимальное кол-во операций
Метод Кабала	2150	2150	2150
Ускоренный метод Кабала	1498	1491	1557
Метод [10]	1309	1066	1936
Предложенный метод	629	428	1428

ния линейных спектральных частот, поиск которых является составной частью большинства современных систем обработки речевых сигналов. В ходе исследования изучена взаимосвязь ЛСЧ с параметрами авторегрессионной (полюсной) модели голосового тракта, а также акустической моделью голосового тракта в виде последовательности труб переменного сечения. Исходя из того, что скорость изменения ЛСЧ определяется динамикой изменения конфигурации голосового тракта, сформулирован вопрос относительно возможности эффективного использования значений ЛСЧ, определенных на некотором временном интервале, для вычисления ЛСЧ следующего временного интервала. В результате проведенного исследования взаимного расположения ЛСЧ на смежных квазистационарных временных интервалах сделаны такие выводы.

1. В преимущественном большинстве ситуаций (от 88.85 % случаев при порядке модели голосового тракта  $p=20$  до 96.78 % случаев при  $p=8$ ) выполняется свойство “межфреймовой упорядоченности” ЛСЧ (5). Это позволяет утверждать, что задача локализации ЛСЧ может быть сведена, главным образом, к простой проверке удовлетворения косинусами ЛСЧ свойства упорядоченности (6). В тех же случаях, когда условие межфреймовой упорядоченности не выполняется, для локализации косинусов ЛСЧ должен использоваться универсальный алгоритм локализации корней трансцендентных уравнений, предложенный в [10, 11].
2. На этапе уточнения значений ЛСЧ для ускорения процесса сходимости могут быть использованы значения косинусов соответствующих ЛСЧ, вычисленные на предыдущем фрейме.

Сделанные выводы легли в основу итогового алгоритма вычисления ЛСЧ, сформулированного в разделе 3. После этого на этапе экспериментальной проверки эффективность предложенного ме-

тода оценивалась при работе с разнообразными тестовыми речевыми сигналами и различными порядками АР модели голосового тракта. В ходе тестирования получены следующие результаты.

1. Итоговая экономия вычислений по сравнению с методом, представленным в [10], не использующим особенности распределения ЛСЧ, составляет от 1.8 до 2.5 раз (при различных порядках полюсной модели голосового тракта).
2. При наиболее часто используемом порядке полюсной модели голосового тракта  $p=10$  предложенный подход к поиску ЛСЧ сокращает вычислительные затраты в 3.4 раза по сравнению с распространенным методом Кабала. Также имеет место существенный выигрыш по сравнению с исследованной “ускоренной” комбинацией метода Кабала и метода Ньютона, равный 2.4 раза.
3. При анализе пиковых значений вычислительных затрат предложенного метода установлено, что его *максимальные* вычислительные затраты являются меньшими не только инвариантных во времени затрат “прямого” метода Кабала, но и *минимальных* затрат ускоренной комбинации метода Кабала с методом Ньютона.

Перечисленные факты свидетельствуют о преимуществе предложенного метода с точки зрения применения в системах реального времени. При этом, поскольку использованные тестовые речевые сигналы отличались разнообразием сменяющих друг друга дикторов и практически полным отсутствием пауз, то приведенные вычислительные характеристики разработанного метода следует воспринимать как соответствующие верхние границы.

1. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов.– М.: Радио и связь, 1981.– 496 с.
2. Маркел Дж., Грей А. Линейное предсказание речи.– М.: Связь, 1977.– 308 с.

3. Full-rate speech transcoding (Recommendation GSM 06.10).– European Telecommunications Standards Institute.– 1992.
4. Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s.– International Telecommunication Union.– 1996.
5. Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction.– International Telecommunication Union.– 1996.
6. Salami R. Design and description of CS-ASELP: a toll quality 8 kbit/s speech coder // IEEE Trans. Speech Audio Proces.– 1998.– 6.– P. 116–128.
7. McCree A. V., Barnwell T. P. A mixed excitation LPC vocoder model for low bit rate speech coding // IEEE Trans. Speech Audio Proces.– 1995.– 4.– P. 242–249.
8. Макхоул Дж., Русос С., Гиш Г. Векторное квантование при кодировании речи // ТИИЭР.– 1985.– 73.– С. 19–61.
9. Калюжный А. Я., Семенов В. Ю. Метод слепой деконволюции речевых сигналов, основанный на анализе линейных спектральных частот // Акуст. вісн.– 2003.– 6, N 3.– С. 34–43.
10. Семенов В. Ю. Новый метод вычисления линейных спектральных частот речевых сигналов, основанный на универсальном алгоритме решения трансцендентных уравнений // Акуст. вісн.– 2002.– 5, N 4.– С. 38–50.
11. Семенов В. Ю. Новый метод локализации корней трансцендентных уравнений // Компьютерна математика.– 2004.– Вып. 1.– С. 143–151.
12. Itakura F. Line spectrum representation of linear predictive coefficients of speech signals // J. Acoust. Soc. Amer.– 1975.– 57, N 1, Suppl. 1.– P. S35.
13. Kabal P., Ramachandran R. P. The computation of line spectral frequencies using Chebyshev polynomials // IEEE Trans. Acoust. Speech Signal Proces.– 1980.– 28.– P. 562–574.
14. Ramachandran R.P., Sondhi M.M., Seshadri N., Atal B.S. A two codebook format for robust quantization of line spectral frequencies // IEEE Trans. Speech Audio Proces.– 1995.– 3, N 3.– P. 157–167.
15. Soong K. S., Juang B.-H. Optimal quantization of LSP parameters // IEEE Trans. Speech Audio Proces.– 1993.– 1.– P. 15–24.
16. Pan J., Fischer T. R. Vector quantization of speech line spectrum pair parameters and reflection coefficients // IEEE Trans. Speech Audio Proces.– 1998.– 6.– P. 106–115.
17. Paliwal K., Atal B. S. Efficient vector quantization of LPC parameters at 24 bits/frame // IEEE Trans. Speech Audio Proces.– 1993.– 1.– P. 3–14.
18. Paliwal K. A study of line spectrum pair frequencies for speech recognition // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– New York, USA, 1988.– P. 485–488.
19. Liu C., Lin M., Wang W., Wang H. A study of line spectrum pair frequencies for speaker recognition // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– Albuquerque, USA, 1990.– P. 277–280.
20. Hansen J., Clements M. Constrained iterative speech enhancement with application to speech recognition // IEEE Trans. Acoust. Speech Signal Proces.– 1991.– 4.– P. 795–805.
21. Kim S.-J., Oh Y.-H. Split vector quantization of LSF parameters with minimum dLSF constraint // IEEE Signal Proces. Lett.– 1999.– 6, N 9.– P. 227–232.
22. Rothweiler J. A rootfinding algorithm on line spectral frequencies // Proc. IEEE Int. Conf. Acoust. Speech Signal Proces.– Phoenix, USA, 1999.– P. 661–664.
23. Rothweiler J. On polynomial reduction in the computation of LSP frequencies // IEEE Trans. Speech Audio Proces.– 1999.– 7.– P. 592–594.
24. Wu C.-H., Chen J.-H. A novel two-level method for the computation of LSP frequencies using a decimation-in-degree algorithm // IEEE Trans. Speech Audio Proces.– 1997.– 5.– P. 106–115.
25. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров.– М.: Наука, 1977.– 834 с.
26. Молчанов И.Н. Машинные методы решения прикладных задач. Алгебра, приближение функций.– К.: Наук. думка, 1987.– 288 с.