

## ВОЗМОЖНОСТЬ ПРИМЕНЕНИЯ СОВРЕМЕННЫХ МЕТОДОВ СЖАТИЯ ИНФОРМАЦИИ ДЛЯ СЖАТИЯ РЕЧИ

### Введение

В современных сетях существенная часть пропускной способности каналов используется для передачи данных и речи. При этом потребности передачи речи значительны. Поэтому важным является эффективное использование полосы пропускания используемые речевыми сигналами. На сегодня передача одной секунды звука с качеством телефонных переговоров требует чуть меньше килобайта данных.

Для обеспечения большей эффективности обмена речевыми сообщениями можно предложить следующие методы:

- 1) Дальнейшее снижение требований к пропускной способности каналов связи при сохранении достигнутого на данный момент качества.
- 2) Достижение качества не ниже уровня аналоговых телефонных каналов при сохранении необходимой скорости передачи.

Целью настоящей статьи является проведение сравнительного анализа методов сжатия речи и определение наиболее перспективных из них с целью применения в IP-телефонии.

### 1. Анализ методов сжатия речи

Частотный диапазон человеческого голоса равен примерно 20-20000 Гц и может значительно отличаться у отдельных людей. Исследования показали, что без существенных потерь диапазон может быть уменьшен до 100-6000 Гц. Но в самом начале строительства телекоммуникаций предоставить такую полосу частот каждому абоненту было невозможно, и пришлось ради удешевления услуг связи пойти на большое сокращение частотного диапазона до полосы 200-3400 Гц. Это привело к ухудшению разборчивости, но она осталась все же на достаточном уровне. В этой полосе частота выборки согласно теореме Котельникова должна составить не менее 8 кГц. Если при этом использовать 8-разрядный код, то необходимая пропускная способность должна составить не менее 64 кбит/с. Обзор современных методов сжатия речи, а также их сравнение было проведено в [1]. За последнее время не было произведено радикальных изменений и бесспорным лидером остался стандарт G.723.1 MP-MLQ. Этот алгоритм является совместной разработкой фирм Audioscodec (Израиль) и DSP Group (США) - кодек с линейным предсказанием и скоростью выходного потока информации 6,3 кбит/с [2]. Несмотря на значительно большую, чем у АДИКМ среднеквадратичную погрешность синтеза речи, алгоритмом получена лучшая (3,9 балла MOS) оценка качества, чем у АДИКМ. Это достигнуто благодаря двум усовершенствованиям алгоритма линейного предсказания алгебраического CELP. Первое: длина сглаживающего окна трехкратно увеличена относительно длины анализируемого сегмента речи, что ослабило искажения, вносимые асинхронностью анализируемых сегментов речи и интервалов основного тона. Второе: метод более точного формирования сигнала возбуждения синтезирующего фильтра, получивший название MP-MLQ (Multipulse Maximum Likelihood Quantisation). Еще одним важным преимуществом этого алгоритма является более надежное, по сравнению с известными аналогами вычисление параметров основного тона голоса. Алгоритм MP-MLQ относится к семейству analysis-by-synthesis алгоритмов. В кодере, построенном на MP-MLQ, используется линейное прогнозирование 10-го порядка, и он может работать на скоростях 4,8; 6,4; 7,2 и 8,0 кбит/с. Гибкость алгоритма дает возможность выполнять адаптацию в ходе разговора при ухудшении связи, а также снизить скорость до 4,0 кбит/с. Для сравнения в табл. 1. приведены характеристики некоторых других алгоритмов.

Таблица 1

Скорость передачи, кбит/с	Субъективное качество	Название стандарта	Год выпуска	Алгоритм	Область применения
64	4,1	ITU-T G.711	1960	PCM	Телефонные сети
32	3,8	ITU-T G.726	1984	ADPCM	Телефонные сети
6,4	3,1	INMAR-SAT-M	1990	IMBE	Спутниковая телефония
13	3,3	ETSI GSM	1992	RPE-LTP	Сотовая телефония (Европа)
16	3,6	ITU-T G.728	1992	LD-CELP	Телефонные сети
4,8	3,4	ETSI TETRA	1996	ACELP	Сотовая телефония (Европа)
6,3	3,9	ITU-T G.723.1	1996	MP-MLQ	Телефонные сети
5,3	3,7	ITU-T G.723.1	1996	ACELP	Телефонные сети
8	3,9	ITU-T G.729	1997	CS-ACELP	Телефонные сети
2,4	3,5	США (проект)	1998	MELP	Мин. Обороны США

Однако, несмотря на то, что прошло достаточно времени с момента утверждения стандарта G.723.1, и на доступность его реализаций, он не имеет той популярности, какую имеет электронная почта. Частично это можно отнести на счет необходимости микрофона (его нужно купить дополнительно) в то время как клавиатура есть у каждого компьютера. Но это не убедительный факт, так как стоимость достаточной для этих целей гарнитуры не превышает стоимости клавиатуры и существенно не удорожает общую стоимость системы. Очевидно, решающим фактором является значительно большая потребность в пропускной способности. По заявленным на данный момент разработкам существенного уменьшения необходимой пропускной способности не предвидется, если не считать разрабатываемого Министерством обороны США стандарта MELP с заявленной скоростью 2,4 кбит/с.

Проведенный анализ показывает, что основные разработки по сжатию речи на протяжении десятков лет ведутся в двух направлениях:

- 1) Разложение в ряд Фурье и применение психоакустической модели для отфильтрования невоспринимаемой человеком информации (MPEG).
- 2) Использование линейного предсказания речи (G.723.1 MP-MLQ).

При этом коэффициент сжатия информации не превышает 20-30. В то же время алгоритмы сжатия графической информации обеспечивают сжатие до 350 раз. Это дает надежду на разработку эффективных алгоритмов сжатия и речи.

## 2. Перспективные методы сжатия речи

Из новых методов сжатия графической информации можно выделить два: с использованием фракталов и с использованием вейвлетов.

Рождение фрактальной геометрии связывают с именем бельгийского математика Мандельброта. В 1977 году он опубликовал книгу «The Fractal Geometry of Nature». Вскоре математики освоили методы построения фрактальных изображений и перешли к вопросу создания системы функций для восстановления заданного изображения. В 1988 году в журнале BYTE Бансли опубликовал статью, в которой приводилась система итерированных функций и получаемое по ним графическое изображение. Коэффициент сжатия при этом составил 10000. Однако, как оказалось, данный алгоритм требовал присутствия человека, а сжатие длилось 100 часов. Полная автоматизация этого процесса снижает коэффициент сжатия до 100.

Преимуществом фрактального метода сжатия является то, что этот метод не анализирует каждый пиксел изображения, поэтому получается лучшее качество изображений любого разрешения. Природа фрактальных преобразований такова, что данные, теряемые при кодировке, наименее существенны для визуального восприятия, например, тонкие детали в текстурах. В отличие от других методов, фрактальное сжатие позволяет контролировать теряемые данные. Наиболее известным из фрактальных форматов файлов является – FIF (Fractal Image Format).

Задача сжатия изображения целиком не реализована и поэтому ее решают поэтапно. В качестве примера можно привести два вида разбиения: Quadtree-разбиение и HV-разбиение.

При Quadtree-разбиении целевой квадратный участок делится на четыре подквадрата в том случае, если он плохо описывается. Процесс продолжается рекурсивно до тех пор, пока ошибка не станет меньше заданной. Слабость метода Quadtree-разбиения в том, что при его реализации нет возможности выбрать исходные элементы контекстно-независимо. Для достижения хороших результатов количество вариантов должно быть очень большим. HV-метод позволяет уменьшить их количество, увеличивая гибкость разбиения. На каждом шаге рассматриваемая прямоугольная область делится на две части вертикальной или горизонтальной линией. Гибкость такой системы выражается в том, что линия разбиения может быть выбрана произвольным образом.

К основным недостаткам фрактальных методов при их применении для сжатия речи можно отнести необходимость накопления больших объемов информации, что приводит к многосекундным задержкам. Это противоречие может быть разрешено с помощью некоторой базы фракталов, аналогичной базе фонем для вокодера.

Другим современным способом сжатия информации является сжатие с использованием вейвлетов[3]. Вейвлеты и вейвлетное преобразование появились как развитие Фурье-преобразования для решения парадокса всегда существующих волн. Вейвлеты так же, как синусы и косинусы, относятся к ортогональным функциям. В остальном вейвлеты отличаются от них. Во-первых, вейвлеты конечны, во-вторых, вместо амплитудного масштабирования в вейвлет-преобразовании используется временная «растяжка», в-третьих, операция сложения в Фурье-преобразовании заменяется временными сдвигами и масштабированием.

Вейвлет-анализ определяется выражением:

$$\gamma(s, \tau) = \int f(t) \Psi_{s, \tau}^*(t) dt, \quad (1)$$

где  $f(t)$  – исходный сигнал, а  $\Psi_{s, \tau}(t)$  – базовая функция, называемая вейвлетом. Вейвлеты получают от одного основного вейвлета  $\Psi(t)$ , называемого материнским, через масштабирование и сдвиги:

$$\Psi_{s, \tau}(t) = \frac{1}{\sqrt{s}} \Psi\left(\frac{t-\tau}{s}\right), \quad (2)$$

где  $s$  – коэффициент масштабирования, а  $\tau$  – временной сдвиг.

Обратное вейвлет-преобразование имеет вид:

$$f(t) = \iint \gamma(s, \tau) \Psi_{s, \tau}(t) dt ds. \quad (3)$$

Из данных выражений следует возможность выбора в качестве материнского вейвлета функции, которая в наибольшей степени удовлетворяет нуждам задачи, если только она удовлетворяет определенным условиям. Одним из таких условий является условие допустимости и непрерывности[4]:

$$\int \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty. \quad (4)$$

Из выражения (4) следует, что вейвлет должен иметь представление, подобное спектру. Так же важно, чтобы среднее значение вейвлета было равно нулю, т.е.

$$\int \Psi(t) dt = 0. \quad (5)$$

Данные преобразования предназначены для аналоговой обработки. Вместе с тем компьютерные системы имеют дело с дискретными представлениями непрерывных сигналов. В выражении (1) вейвлетное преобразование вычисляется непрерывным сдвигом и масштабированием функции и вычислением корреляции между ними. Но основной проблемой является невозможность аналитически вычислить результат вейвлет-преобразования и их вычисляют численными методами. Из всего вышесказанного следует, что в дискретном преобразовании сигнал масштабируется и сдвигается не непрерывно, а только с определенным дискретным шагом:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right), \quad (6)$$

где  $j$  и  $k$  – целые числа и  $s_0 > 1$  – фиксированный шаг. Коэффициент сдвига  $\tau_0$  определяется шагом. Обычно  $s_0$  выбирается равным 2 как наиболее естественный для компьютерных реализаций. Значение для  $\tau_0$  обычно выбирается равным 1.

Когда дискретные вейвлеты используются для преобразования непрерывного сигнала, результатом будет серия коэффициентов вейвлета, называемая последовательной декомпозицией вейвлета. Важным в схеме декомпозиции есть вопрос восстановления. В [5] доказывается, что необходимое и достаточное условие для восстановления состоит в том, что энергия коэффициентов вейвлета должна лежать между двумя положительными границами, то есть:

$$A \|f\|^2 \leq \sum_{j,k} \left| \langle f, \Psi_{j,k} \rangle \right|^2 \leq B \|f\|^2, \quad (7)$$

где  $\|f\|^2$  – энергия сигнала  $f(t)$ ;  $A > 0$ ;  $B < \infty$  и  $A, B$  – независимы от  $f(t)$ . Когда условие (7) выполняется, семейство функций  $\Psi_{j,k}(t)$  с  $j, k \in Z$  называется фреймом, а  $A$  и  $B$  – границами фрейма. Когда  $A = B$ , фрейм сжат и дискретные вейвлеты ведут себя подобно ортонормальной основе. Когда  $A \neq B$ , точное восстановление возможно за счет двойного фрейма. В двойном фрейме дискретного вейвлета преобразующий вейвлет отличается от восстанавливающего вейвлета.

На последнем шаге нужно сделать дискретные вейвлеты ортогональными. Дискретные вейвлеты могут быть сделаны ортогональными в своих растяжениях и сдвигах специальным выбором «материнского» вейвлета:

$$\int \Psi_{j,k}(t) \Psi_{m,n}^*(t) dt = \begin{cases} 1, & \text{если } j = m \text{ и } k = n \\ 0, & \text{в других случаях} \end{cases}. \quad (8)$$

Произвольный сигнал может восстанавливаться суммированием ортогональных базисных функций, с весом коэффициентов вейвлетного преобразования:

$$f(t) = \sum_{j,k} \gamma(j,k) \Psi_{j,k}(t) \quad (9)$$

Ортогональность не всегда необходима для представления сигналов. В некоторых приложениях могут использоваться неортогональные вейвлеты. При этом избыточность может

уменьшать чувствительность к шуму [4]. К недостатку дискретных вейвлетов можно отнести следующее: результирующее вейвлет-преобразование больше не инвариантно к сдвигу, который означает, что вейвлет-преобразование сигнала и сдвинутого во времени того же сигнала, не являются просто сдвинутыми версиями друг друга.

Чтобы сжать с помощью вейвлетов информацию, необходимо уменьшить количество нужных для восстановления вейвлетов. Даже с дискретными вейвлетами нужно большое количество масштабирований и сдвигов, чтобы вычислить вейвлет-преобразование. Для упрощения задачи можно отказаться от использования бесконечного количества дискретных вейвлетов. Исследования показывают, что от выбора количества вейвлетов зависит и коэффициент сжатия и качество восстанавливаемой информации. Сдвиги вейвлетов являются ограниченными длительностью сигнала, и это есть верхняя граница для вейвлета. Но данное обстоятельство не снимает вопроса о необходимом количестве масштабирований для анализа сигнала. Выше было указано, что вейвлеты имеют полосу частот подобную спектру. Из теории Фурье известно, что временное сжатие эквивалентно сжатию спектра и смещению его в верхнюю область частот:

$$F\{f(at)\} = \frac{1}{|a|} F\left(\frac{\omega}{a}\right). \quad (10)$$

Это означает, что временное сжатие вейвлета в 2 раза приведет к сжатию частотного спектра вейвлета в 2 раза, а также сдвинет все частоты в 2 раза в верхнюю область. Указанное предопределяет возможность перекрытия конечного спектра нашего сигнала со спектрами растянутых и сдвинутых вейвлетов. Для более точного описания сигнального спектра, растянутые спектры вейвлетов должны касаться друг друга. Результат масштабирования материнского вейвлета показан на рис. 1.

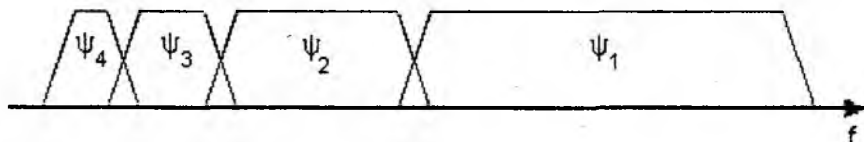


Рис. 1

Из этого следует, что если один вейвлет может быть представлен как частотный фильтр, то серия расширенных вейвлетов может быть представлена как набор фильтров частот. Если вычислить коэффициент, определяющий отношение между центральной частотой спектра вейвлета и шириной этого спектра, то он будет одинаков для всех вейвлетов. Этот коэффициент называют показателем точности Q-фильтра. В случае вейвлетов он называется константой Q-банка фильтров.

Если представить вейвлет-преобразование как группу фильтров, то можно рассматривать вейвлетное преобразование сигнала как прохождение сигнала через эту группу фильтров. Выходами с различных ступеней фильтра являются вейвлет и функция масштабирования превращающих коэффициентов. Следует указать, что анализ сигнала, прошедшего через банк фильтров, идея не новая и используется в течение многих лет в частотном кодировании.

Банк фильтров, необходимый для частотного кодирования, может быть создан различными путями. Одним из них является построение большого числа фильтров частотных полос, чтобы разделить спектр на различные полосы. Преимущество этого метода состоит в том, что ширина каждой полосы может быть выбрана произвольно. Недостатком метода является то, что в этом случае необходимо создавать каждый фильтр отдельно, а это может быть трудоемким процессом. Другой путь состоит в том, чтобы разделить сигнальный спектр на две равные части – низкочастотную и высокочастотную. Высокочастотная часть содержит индивидуальные характеристики и особенности сигнала, которые должны быть сохранены. Тем не менее, низкочастотная часть все еще содержит некоторые особенности

сигнала и, следовательно, мы можем разделить ее снова. Так происходит до тех пор, пока не достигнуто требуемое количество полос. Обычно количество полос ограничивается длиной данных или доступной вычислительной мощностью процессора системы. Процедура разбиения спектра представлена на рис. 2. Преимущество этой схемы в том, что необходимо создать только два фильтра.

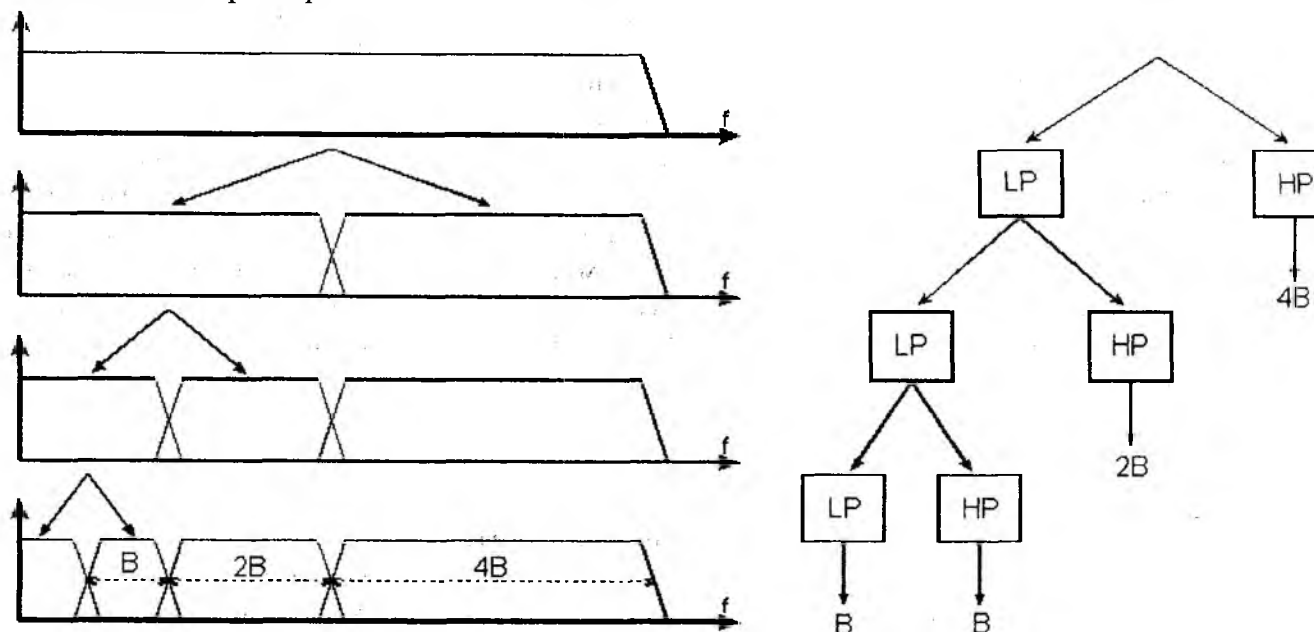


Рис. 2

Фирма Analog Devices выпускает недорогую микросхему ADV601, предназначенную для сжатия видеосигнала с коэффициентом сжатия 350:1 [6]. В ней используется биортогональное (7,9) вейвлет-преобразование и аналогичный рассмотренному выше метод разбиения. Различие состоит в том, что вейвлет-преобразование используется для обработки двухмерного сигнала, и это накладывает свои особенности. Также компания AT&T разработала графический формат хранения информации под названием DJVU.

### Заключение

К настоящему времени разработаны и используются различные математические методы обработки и представления сигналов, в том числе и с использованием вейвлетов. На их основе созданы аппаратные и программные реализации применительно к сжатию и обработке графической информации. При этом использование этих методов для приложений, связанных с обработкой речевой информации, ограничено. Из указанного следует необходимость проведения дальнейшего исследования о применении теории вейвлетных преобразований в качестве возможной альтернативы существующим методам.

**Список литературы:** 1. Замула А.А., Павленко Ю.С. Защита информации в IP-телефонии// Радиотехника: Всеукр. межвед. науч.-техн. сб. 2001. Вып.119. С.191-194 2. ITU-T Recommendation G.723.1 Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s. 3. Valens, C. A Really Friendly Guide to Wavelets, 1999. 4. Sheng, Y. Wavelet Transform, 1996. 5. Daubechies, I. Ten Lectures on Wavelets. 2nd ed. Philadelphia: SIAM, 1992. 6. Analog Devices, Inc, 1997. ADV601 Low Cost Multiformat Video Codec.

Харьковский национальный  
университет радиоэлектроники

Поступила в редколлегию 15.04.2002