

Использование нелинейных ортогональных преобразований

для сжатия сигналов звукового диапазона

Григоров И.В.

Рассматриваются существующие методы сжатия сигналов и изображений с использованием дискретных ортогональных преобразований. Описывается возможность использования для этих же целей нелинейных унитарных операторов, в частности нелинейного дифференциального оператора Шредингера. Рассматриваются вопросы цифровой реализации этих операторов с использованием нелинейных фазовых фильтров. Обсуждаются вопросы оптимизации этих фильтров и их реализации с использованием математического пакета «MathCad». Приводятся результаты исследования алгоритмов сжатия звуковых сигналов.

Для сжатия непрерывных сигналов и изображений, представленных в цифровой форме, часто используются различные ортогональные преобразования. Чаще всего для этих целей используются дискретное преобразование Фурье (ДПФ) в форме алгоритма быстрого преобразования Фурье (БПФ), дискретное косинусное преобразование (ДКП), различные вейвлет-преобразования и т.д. [1,2]. Все эти преобразования по определению являются линейными. Идея всех этих методов состоит в следующем. Исходный вектор сигнала \mathbf{s} размерности N подвергается линейному преобразованию \mathbf{F} . Этим осуществляется переход в некоторую обобщенную спектральную область. Преобразование выбирают таким, чтобы число компонент M сигнала в этой области, имеющих существенные значения, т.е. превышающих некоторое пороговое значение P , было минимальным. Описанное свойство оператора \mathbf{F} можно назвать преселектирующим. Остальные компоненты спектра, число которых $(N - M)$, в простейшем случае отбрасывают. Эту операцию обычно называют пороговой амплитудной селекцией. Она является нелинейной и может быть описана нелинейным оператором \mathbf{K} . Алгоритм амплитудной селекции, который может быть и более сложным, определяет степень сжатия сигнала. В простейшем случае пороговой селекции степень сжатия определяется выбором величины порога селекции P . Чем он выше, тем больше степень сжатия и тем больше погрешность его восстановления. Количественно степень сжатия можно оценить коэффициентом сжатия [1]:

$$Q = \frac{N}{M}. \quad (1)$$

Восстановление сигнала производится с помощью обратного преобразования \mathbf{F}^{-1} . Таким образом, описанную процедуру сжатия - восстановления можно представить в операторной форме следующим образом:

$$\hat{\mathbf{s}} = \mathbf{F}^{-1} \mathbf{K} \mathbf{F} \mathbf{s}, \quad (2)$$

где \mathbf{s} и $\hat{\mathbf{s}}$ – векторы исходного и восстановленного сигналов, \mathbf{F} и \mathbf{F}^{-1} – операторы прямого и обратного ортогональных преобразований, \mathbf{K} – нелинейный оператор амплитудной селекции компонент спектра.

Качество восстановления сигнала чаще всего оценивают по величине среднего квадрата ошибки (СКО):

$$\overline{\varepsilon^2} = \overline{\|\hat{\mathbf{s}} - \mathbf{s}\|^2}, \quad (3)$$

где черта сверху означает знак усреднения по множеству реализаций. Очевидно, что с увеличением коэффициента Q растет величина СКО. Процедуру (2) можно оптимизировать по критерию минимума величины (3) путем выбора оператора \mathbf{F} , обладающего наилучшими преселектирующими свойствами, и выбора нелинейного оператора амплитудной селекции \mathbf{K} .

Иногда, в частности, при сжатии звуковых сигналов с помощью стандартных кодеков семейства MPEG [3], обработка сигналов производится в частотной области. При этом в качестве оператора \mathbf{F} используют алгоритм модифицированного ДКП (МДКП), который строится на основе алгоритма БПФ. При реализации обратного оператора \mathbf{F}^{-1} используются алгоритмы обратного МДКП и обратного БПФ. С целью повышения степени сжатия, процедуру амплитудной селекции в частотной области усложняют и пренебрегают не

только слабыми спектральными компонентами, но и теми, которые расположены близко по частоте к сильным. Это позволяет повысить коэффициент сжатия Q . Для дальнейшего повышения коэффициента сжатия описанную процедуру усложняют, используя коды Хаффмена. На сегодняшний день алгоритмы стандартов MPEG являются самыми эффективными и позволяют достичь значений величины g порядка 10-11 для высококачественных музыкальных сигналов и 30-50 для речевых сигналов с приемлемым на слух качеством восстановления.

К недостаткам алгоритмов MPEG следует отнести их сложность, а также относительно большое значение величины СКО восстановления (3). Например, при значении коэффициента $Q \approx 11$, что имеет место при кодировании двухканальных музыкальных сигналов формата «WAV» в формат «MPEG-1 ISO/IEC 11172-3» с уровнем компрессии «Layer 3» (скоростью цифрового потока – «BitRate» – 128 кБит/с), отношение средней мощности полезного сигнала к величине (3) для различных сигналов имеет значение порядка 12–20 дБ. Хотя это и не ведет к заметному искажению сигнала, что объясняется эффектом маскировки слуха [3], алгоритмы MPEG в чистом виде практически нельзя использовать в тех случаях, когда требуется максимально точно восстановить форму сигнала, т.е. минимизировать СКО (3).

Для решения задачи сжатия сигналов по критерию минимума величины (3) в отличие от методов, использующих линейные ортогональные преобразования, предлагается в качестве F использовать оператор с унитарной нелинейностью, описывающий нелинейный фильтр с распределенными параметрами. Такие операторы используются для описания эволюции нелинейных волн в различных диспергирующих средах, в частности рассматриваются в нелинейной квантовой механике [4], нелинейной оптике [5]. Кроме того, они могут быть применены для селективного подавления импульсных и сосредоточенных помех. Эти методы, описанные подробнее в [6,7], основаны на преселектирующих свойствах операторов с

унитарной нелинейностью, которыми обладают упомянутые выше линейные ортогональные преобразования.

Нелинейный дифференциальный оператор, описывающий в частности эволюцию огибающей оптического импульса в оптическом волокне без затухания при больших интенсивностях импульсов, можно представить нелинейным уравнением Шредингера [5]

$$i \frac{\partial \Psi}{\partial \eta} + \alpha \frac{\partial^2 \Psi}{\partial \tau^2} + f(\Psi)\Psi = 0, \quad (4)$$

где $\Psi(\eta, \tau)$ – комплексная огибающая импульса, распространяющегося вдоль пространственной координаты η ; τ – «бегущая» временная координата импульса, отсчитываемая от его центра, α – коэффициент фазы, $f(\Psi)$ – функция, определяющая вид нелинейности оптического волокна (в простейшем случае это квадратичная функция, а соответствующая нелинейность называется кубической). Огибающую $\Psi(\eta, \tau)$ в начале волокна, т.е. при $\eta = 0$, можно рассматривать как сигнал на входе эквивалентного нелинейного фильтра (описываемого оператором F) а при значении координаты $\eta = l$, т.е. $\Psi(l, \tau)$, – выходным сигналом этого фильтра:

$$\Psi(0, \tau) = u(\tau); \quad \Psi(l, \tau) = v(\tau). \quad (5)$$

Для практической реализации нелинейного фильтра в цифровой форме целесообразно использовать метод расщепления по физическим факторам [8]. Он используется для численного решения волновых уравнений и моделирования эволюции огибающей оптических импульсов вдоль оптических волокон. Он заключается в том, что участок волокна длиной l разбивается на n участков длиной $\Delta\eta$. При этом, считается, что между участками действует только нелинейность, а на протяжении каждого участка – только дисперсия. Таким образом, нелинейный оператор F можно представить, в виде произведения нелинейных (H_k) и линейных (G_k) операторов вида:

$$F(\Psi) = \prod_{k=1}^n G_k H_k(\Psi). \quad (6)$$

Каждый из линейных операторов \mathbf{G}_k может быть представлен в виде линейного фазового фильтра, реализующего преобразование Френеля во временной области. Его импульсная характеристика (функция Грина) и соответствующая передаточная функция имеют вид:

$$g(\tau) = g_0 \exp\left(i \frac{a\tau^2}{2}\right), \quad (7)$$

$$G(i\omega) = \exp(-i\alpha\Delta\eta\omega^2), \quad (8)$$

где a и g_0 - постоянные коэффициенты:

$$a = \frac{1}{2\alpha\Delta\eta}, \quad g_0 = \frac{1}{\sqrt{4\pi\alpha\Delta\eta}} \exp\left(i \frac{3\pi}{4}\right). \quad (9)$$

Нелинейный оператор \mathbf{H}_k может быть реализован как нелинейный преобразователь фазы с характеристикой преобразования мгновенных значений

$$H(u) = \exp(iff(u)). \quad (10)$$

Обратный оператор \mathbf{F}^{-1} описывается сопряженным уравнением

$$-i \frac{\partial \Psi}{\partial \eta} + \alpha \frac{\partial^2 \Psi}{\partial \tau^2} - f(\Psi)\Psi = 0. \quad (11)$$

При его реализации также можно использовать аналогичные линейные и нелинейные звенья с характеристиками, сопряженными с (7), (8) и (10), расположив их в обратном порядке, т.е.

$$\mathbf{F}^{-1}(\Psi) = \mathbf{F}^*(\Psi) = \prod_{k=1}^n \mathbf{H}^*(\Psi) \mathbf{G}_k^*. \quad (12)$$

Поскольку все рассмотренные звенья имеют единичные коэффициенты передачи по амплитуде, рассмотренные фильтры называются нелинейными фазовыми фильтрами (НФФ).

Как показано в [6] для подавления импульсных помех в каналах связи можно использовать простейший однозвенный НФФ, содержащий одну пару звеньев, описываемых операторами \mathbf{H} и \mathbf{G} с характеристиками (10) и (8). Такой НФФ при надлежащем выборе нелинейной функции $f(\Psi)$ и коэффициента фазы α обладает преселектирующим свойством. Это проявляется в том, что при прохождении через такой НФФ импульсный сигнал гладкой формы сжимается во времени и растет по амплитуде, что аналогично

эффекту аномальной дисперсии в оптических волокнах. За счет этого происходит перераспределение мощности между отсчетами сигнала – мощность «сильных» отсчетов увеличивается, а «слабых» – уменьшается. При этом число «сильных» отсчетов уменьшается, а «слабых» – растет. Этот эффект и предлагается использовать для сжатия непрерывных сигналов.

Простейший алгоритм сжатия – восстановления сигнала с использованием однозвенного НФФ имеет вид, приведенный на рисунке 1.

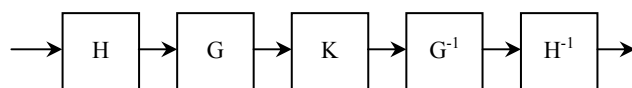


Рис.1.

Данный алгоритм был смоделирован на компьютере с использованием математического пакета «MathCad – 2003». В качестве входного сигнала, подлежащего сжатию, использовался звуковой сигнал формата «Wav». Это позволяет исследовать предлагаемый алгоритм как для обработки стандартных тестовых сигналов (синусоидальных, импульсных, шумовых и т.д.), так и для обработки реальных сигналов, которые можно записать в компьютер с помощью звуковой платы или других внешних устройств. Качество восстановления сигналов при этом можно оценивать не только количественно по величине (3), но и качественно, путем прослушивания и сравнения исходного и восстановленного сигналов.

При реализации нелинейного преобразователя фазы с характеристикой преобразования (10) нелинейная функция $f(u)$ задавалась в полиномиальной форме

$$f(u) = \sum_{k=0}^{N-1} \beta_k |u|^k, \quad (13)$$

где β_k – вектор неизвестных коэффициентов полинома, подлежащих оптимизации, N – их число. Вектор оптимальных параметров был найден вариационным методом, описанным в [6], в соответствии с критерием максимума пиковой мощности сигнала на выходе НФФ.

Линейный фильтр Френеля имеет постоянный коэффициент передачи по амплитуде и квадратичную фазовую характеристику (см. (8)). Его импульсная характеристика (7) является комплексной функцией времени бесконечной длины. При реализации фильтра в цифровой форме ее нужно ограничить с помощью окна. Здесь было использовано модифицированное окна Ханна [2]. Действительная и мнимая части полученной импульсной характеристики фильтра имеют вид, представленный на рисунках 2 и 3:

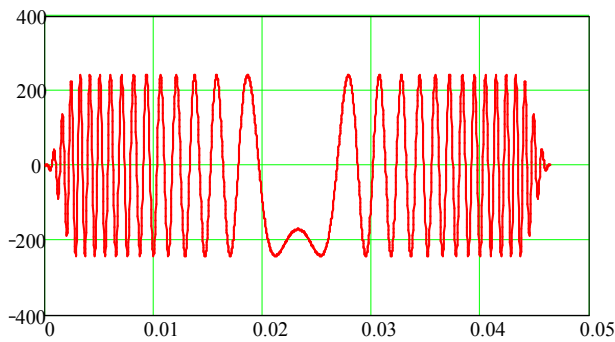


Рис.2.

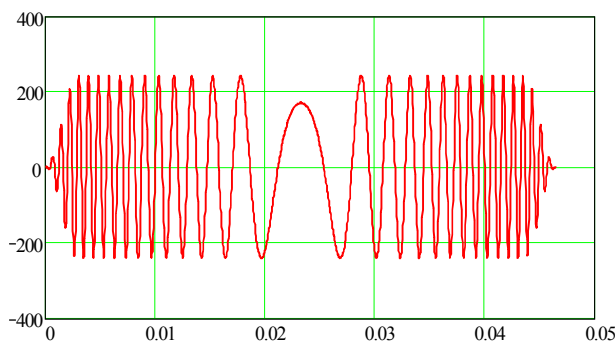


Рис.3.

Коэффициент фазы фильтра α определяет скорость изменения мгновенной частоты импульсной характеристики (при фиксированном параметре $\Delta\eta$, который принимался равным единице). В процессе моделирования он подбирался экспериментально исходя из требований минимальных амплитудных искажений сигнала и максимально широкой полосы пропускания фильтра ΔF . Последняя при правильном выборе α должна быть близкой к половине частоты дискретизации f_d . При моделировании частота дискретизации определялась исходным WAV-сигналом и имела стандартное значение $f_d = 44,1$ кГц. При этом была получена полоса пропускания фильтра $\Delta F \approx 19$ кГц. АЧХ фильтра приведена на рисунке 4.

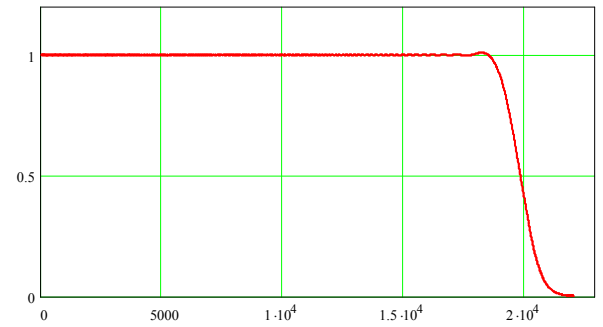


Рис.4.

В качестве оператора \mathbf{K} при проведении экспериментов использовался простейший алгоритм пороговой амплитудной селекции. Исследуемый звуковой сигнал представлял собой фрагмент фонограммы «Flash Dance» группы «Deep Dish» длительностью 3 секунды. Некоторые результаты экспериментов приведены на рисунках 5 и 6. На них представлены соответственно зависимости коэффициента сжатия Q и отношения средней мощности исходного сигнала к СКО (в децибелах) от порога селекции P . Его величина задается в единицах номеров уровней сигнала и может меняться для данной разрядности сигнала ($n_p = 16$) от 0 до 32767.

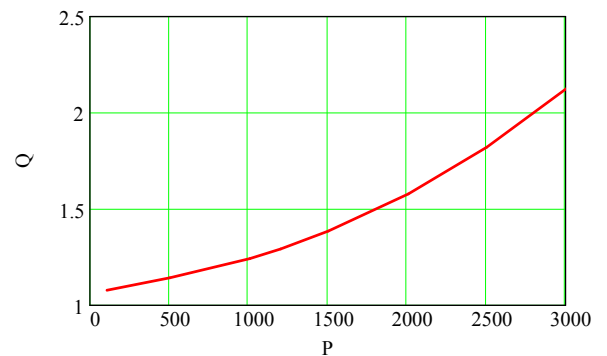


Рис. 5.

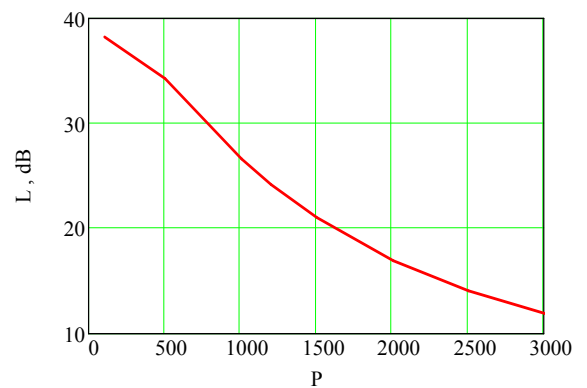


Рис. 6.

Очевидно, что с увеличением степени сжатия сигнала повышается и СКО, т.е. уменьшается

отношение L . Нужно отметить, что при достаточно существенном коэффициенте сжатия $Q \approx 1.4$ (или 40% от исходного объема) обеспечивается отношение $L \approx 21$ дБ. Качество восстановления музыкального сигнала при этом довольно высокое, т.к. звучание восстановленной фонограммы почти не отличается от исходной. На рисунках 7 и 8 приведены временные диаграммы этих сигналов для указанной величины Q . Задержка второго сигнала, относительно первого составляет приблизительно 46 мс.

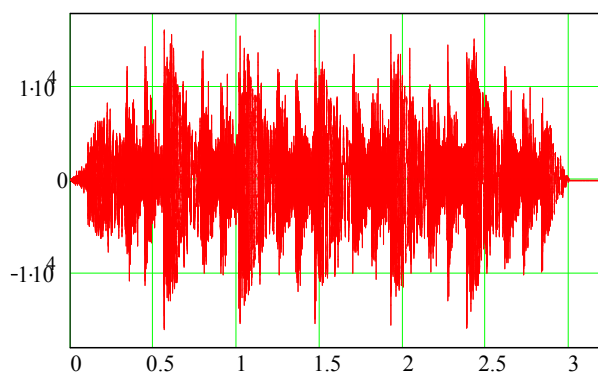


Рис.7.

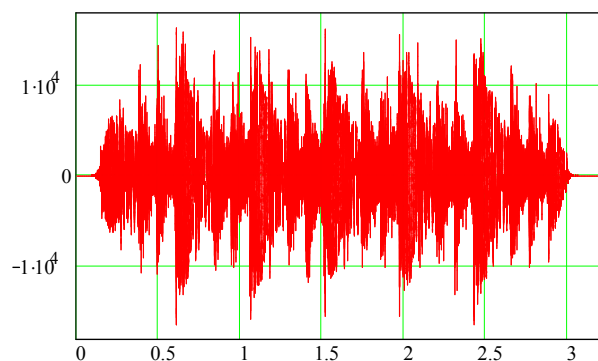


Рис.8.

ЛИТЕРАТУРА

1. Ахмед Н., Рао К.Р. Ортогональные преобразования при обработке цифровых сигналов. – М.: Связь, 1980.– 248 с.
2. Залманзон Л.А. Преобразования Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. – М.: Наука, 1989. – 496 с.
3. Ю.А.Ковалгин, Э.И.Вологдин. Цифровое кодирование звуковых сигналов. – Санкт-Петербург: Корона принт, 2004. – 235 с.
4. В.П.Маслов. Комплексные марковские цепи и континуальный интеграл Фейнмана. – М.: Наука, 1976. – 192 с.
5. С.А.Ахманов, В.А.Выслоух, А.С.Чиркин. Оптика фемтосекундных лазерных импульсов. – М.: Наука, 1988 – 312 с.
6. Shirokov S.M., Grigorov I.V. Supression of impulsive noise at space-time signals and images processing with use of nonlinear phase filters. Proceedings of World Multiconference on Systemics, Cybernetics and Informatics, vol. XII. – Orlando, Florida, USA, 2001, pp. 401-404.
7. Shirokov S.M., Petrov A.V. Digital signal processing using nonlinear orthogonal transforms in frequency domain // Proc. of SCI2002/ISAS2002, vol.14, Orlando, 2002, pp.267–270.
8. Г.Агравал. Нелинейная волоконная оптика. – М.: Мир, 1996. 328 с.

Use of non-linear orthogonal transformation for compressing of sound signals

Grigorov I.V.

Existing methods of compression of signals and images with use of discrete orthogonal transformations are considered. The opportunity of use for the same purposes of nonlinear unitary operators, in particular nonlinear differential operator of Schroedinger is described. Questions of realization of these operators with use of nonlinear phase filters of Fresnel in the form of digital processing signals are considered. And their realizations questions of optimization of these filters are discussed with use of a mathematical package "MathCad". Results of research of algorithms of compression of sound signals are resulted.
