

К. В. ЮРКОВ, С. Е. ПЕТРОВ

О ДВУХ ПОДХОДАХ К ПАРАМЕТРИЧЕСКОМУ КОДИРОВАНИЮ СТЕРЕОСИГНАЛА

Рассматриваются два подхода к кодированию стереофонического сигнала. Предлагаемые подходы позволяют обеспечить высокое качество синтезированного стереофонического аудиосигнала при небольших затратах на кодирование.

Ключевые слова: параметрический стереосигнал, кодирование аудиосигнала.

Введение. Одна из важных задач обработки цифровых сигналов — кодирование аудиосигналов. В настоящее время существует множество стандартов сжатия монофонических аудиосигналов, доминирующим является алгоритм AAC+ (Advanced Audio Coding) [1].

При постановке задачи кодирования стереосигнала предполагается, что монофонический кодер уже реализован. Известны три основных алгоритма обработки стереофонического сигнала. Кратко остановимся на каждом из них.

Mid-Side Stereo [1]. Данный алгоритм подразумевает отдельную передачу суммы и разности правого и левого каналов. Понятно, что если полученные таким образом сигналы переданы идеально точно, то этот метод позволяет идеально точно восстановить исходный стереофонический сигнал. Существенным недостатком данного подхода, является необходимость в большинстве случаев передавать два моносигнала, обладающие практически одинаковой энергией.

Intensity Stereo [2]. Данный алгоритм предполагает передачу одного моноканала, полученного как полусумма правого и левого каналов, и набора параметров, определяющих соотношения энергий правого и левого каналов.

Parametric Stereo [3]. Данный алгоритм предполагает передачу одного моноканала, полученного как полусумма правого и левого каналов, и набора параметров. Параметры в этом случае определяют как соотношения энергий правого и левого каналов, так и углы поворота на комплексной плоскости.

Последние два подхода требуют меньших затрат на кодирование (объем информации, в битах, на выходе кодера) по сравнению с алгоритмом “mid-side stereo”, однако существенным их недостатком является невозможность в некоторых случаях восстановить исходную стереопару с достаточным качеством.

Перечисленные алгоритмы имеют как достоинства, так и существенные недостатки. Таким образом, разработка методов кодирования стереофонического аудиосигнала, позволяющих достичь высокого качества его восстановления при малых затратах на кодирование, является актуальной задачей.

На рис. 1 представлена общая схема кодирования стереосигнала на основе спектрального преобразования.

Входной стереосигнал поступает в модуль комплексного спектрального преобразования. В качестве такого преобразования могут быть, в частности, использованы преобразование Фурье, квадратурное зеркальное преобразование [4], модифицированное косинусное преобразование [5] или любой другой банк ортогональных фильтров. Полученные спектры правого (R) и левого (L) каналов поступают в модуль совмещения каналов и извлечения параметров. На выходе данного модуля наблюдается моноспектр, который обрабатывается любым монокодером, и определяются параметры стереосигнала (далее — стереопараметры), которые

также подвергаются кодированию. Кроме того, возможно дополнительное кодирование разностного сигнала.

В настоящей статье рассматриваются два подхода к кодированию стереосигнала. Использование первого подхода не предусматривает кодирования разностного сигнала, а следовательно, не позволяет в общем случае восстановить оригинальный стереосигнал. Второй подход использует кодирование разностного сигнала, но при этом требует дополнительного объема информации для его передачи.

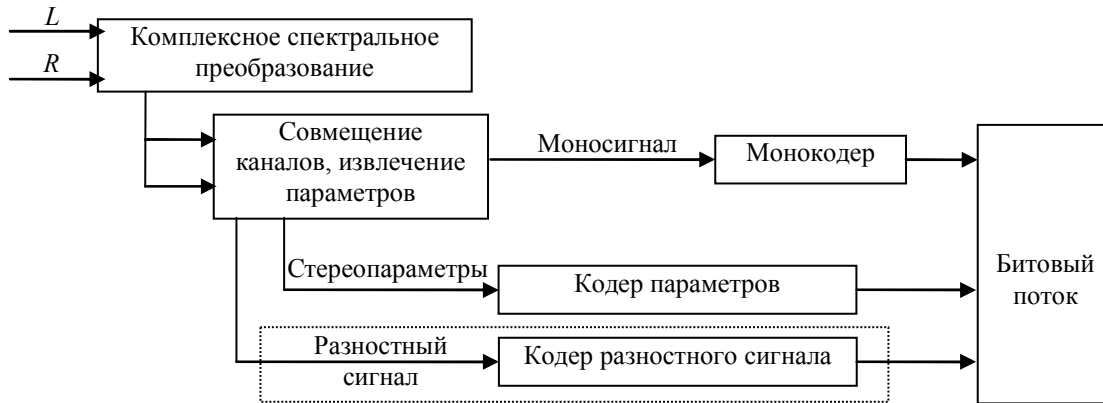


Рис. 1

Совмещение стереоканалов. В случае когда разностный сигнал не передается, возникает проблема получения одного моноканала из двух стереоканалов. Рассмотрим следующую схему.

Введем необходимые обозначения. Обозначим через $\mathbf{l} = (l_1, \dots, l_m) \in \mathbb{R}^m$ и $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^m$ временные кадры длиной m в левом и правом каналах. Через $\mathbf{L} = (L_1, \dots, L_N) \in \mathbb{C}^N$ и $\mathbf{R} = (R_1, \dots, R_N) \in \mathbb{C}^N$ обозначим спектры длиной N соответствующих временных сигналов.

Комплексные значения коэффициентов спектра представим в экспоненциальной форме:

$$L_k = a_k^L \exp(j\varphi_k^L), R_k = a_k^R \exp(j\varphi_k^R), k = 1, \dots, N,$$

где a_k^L и a_k^R — значения амплитуд сигналов в левом и правом каналах, φ_k^L и φ_k^R — соответствующие значения фаз этих сигналов.

Обозначим через $M_k = a_k^M \exp(j\varphi_k^M)$ соответствующее значение совмещенного сигнала. Естественным решением задачи совмещения стереоканалов может служить сигнал

$$Z_k = a_k^Z \exp(j\varphi_k^Z) = \frac{L_k + R_k}{2}.$$

Данное решение является удовлетворительным за исключением варианта, когда сигналы L_k и R_k близки к противоположным, т.е. $L_k \approx -R_k$. При этом амплитуда сигнала Z_k мала, а угол φ_k^Z фактически принимает случайные значения. Для того чтобы избежать ослабления совмещенного сигнала M_k , выберем значение a_k^M как

$$a_k^M = \frac{a_k^L + a_k^R}{2}.$$

Для решения проблемы случайных значений угла φ_k^Z при возникновении противоположных сигналов L_k и R_k вычислим значение φ_k^M согласно формуле

$$\varphi_k^M = \begin{cases} \angle(L_k e^{j\psi} + R_k e^{-j\psi}), a_k^Z / a_k^M < 10^{-3}; \\ \varphi_k^Z, a_k^Z / a_k^M \geq 10^{-3}, \end{cases}$$

где величина поворачивающего угла ψ определена эмпирически и равна $\psi = \pi/100$.

Полученный сигнал M_k используем для оценки параметров исходного стереосигнала.

Оценка параметров стереосигнала. Полагаем, что спектр стереосигнала и спектр совмещенного моно сигнала разделены на подполосы. Параметры амплитуд и фаз сигналов вычисляются для каждой из подполос и передаются в закодированный поток. Через индексы n, k обозначим принадлежность сигнала к компоненту спектра, имеющему номер k в подполосе с номером n .

На рис. 2 приведено графическое представление на комплексной плоскости левого канала ($L_{n,k}$), правого канала ($R_{n,k}$), их оценок ($\hat{L}_{n,k}$, $\hat{R}_{n,k}$) и моноканала ($M_{n,k}$). Как следует из анализа представленного рисунка, значения $\hat{L}_{n,k}$, $\hat{R}_{n,k}$ могут быть получены путем преобразования моно сигнала $M_{n,k}$ при помощи поворота и масштабирования.

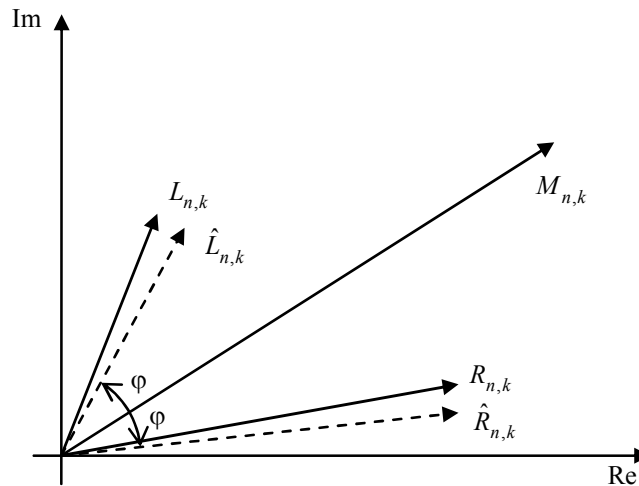


Рис. 2

Согласно стандартной модели образования стереофонического сигнала один и тот же сигнал поступает на два записывающих устройства. Следовательно, левый и правый каналы отличаются лишь задержкой (сдвигом фаз в спектральной области) и коэффициентами усиления.

Так как в качестве амплитуды совмещенного сигнала было выбрано среднее значение амплитуд сигналов правого и левого каналов, естественной формой их восстановления являются оценки

$$\hat{L}_{n,k} = g e^{j\varphi} M_{n,k}, \quad \hat{R}_{n,k} = (2 - g) e^{-j\varphi} M_{n,k}.$$

Таким образом, параметрами, описывающими стереосигнал в подполосе, является пара (g, φ) , при этом $\varphi \in [-\pi, \pi]$, $g \in [0, 2]$.

Иногда, в зависимости от типа спектрального преобразования, для уменьшения затрат на кодирование необходимо вычислять значения стереопараметров одновременно для подполос в группе спектров. Обозначим количество спектров группы через v .

Ошибка восстановления стереосигнала в полосе длиной q может быть вычислена как

$$E(g, \varphi) = \sum_{n=1}^v \sum_{k=1}^q \left(|L_{n,k} - \hat{L}_{n,k}|^2 + |R_{n,k} - \hat{R}_{n,k}|^2 \right).$$

Для выбора параметров стереосигнала необходимо решить следующую экстремальную задачу:

$$E(g, \varphi) \rightarrow \min_{g, \varphi}.$$

Вычислим частные производные функции $E(g, \varphi)$ и приравняем их к нулю. Уравнение

$$\frac{\partial E(g, \varphi)}{\partial g} = 0$$

имеет решение

$$g = 1 + \frac{\sum_{n=1}^v \sum_{k=1}^q a_{n,k}^M a_{n,k}^L - \sum_{n=1}^v \sum_{k=1}^q a_{n,k}^M a_{n,k}^R}{2 \sum_{n=1}^v \sum_{k=1}^q (a_{n,k}^M)^2}.$$

Из уравнения

$$\frac{\partial E(g, \varphi)}{\partial \varphi} = 0$$

находим значение тангенса фазы:

$$\operatorname{tg} \varphi = \frac{\sum_{n=1}^v \sum_{k=1}^q (a_{n,k}^R)^2 \sin(\varphi_{n,k}^M - \varphi_{n,k}^R) + \sum_{n=1}^v \sum_{k=1}^q (a_{n,k}^L)^2 \sin(\varphi_{n,k}^L - \varphi_{n,k}^M)}{\sum_{n=1}^v \sum_{k=1}^q (a_{n,k}^R)^2 \cos(\varphi_{n,k}^M - \varphi_{n,k}^R) + \sum_{n=1}^v \sum_{k=1}^q (a_{n,k}^L)^2 \cos(\varphi_{n,k}^L - \varphi_{n,k}^M)}.$$

Полученные параметры (g, φ) квантуются и передаются в закодированный поток.

Отметим, что аналогичная методика оценки параметров стереосигнала используется в работе [2], однако разница заключается в том, что предлагаемый в настоящей статье подход подразумевает решение задачи минимизации ошибки восстановления.

Оценка параметров стереосигнала при передаче разностного сигнала. Рассмотрим ситуацию, когда необходимо получить идеальное качество восстановления стереосигнала. В этом случае передачи только стереопараметров бывает недостаточно. Однако в рассмотренной выше схеме исправления ошибок, полученных при восстановлении стереосигналов, необходимо передавать два разностных сигнала — для левого и правого каналов. В данной статье предлагается подход, при котором требуется дополнительно передать лишь один разностный сигнал.

Особенность предлагаемого подхода заключается в том, что оценка параметров стереосигнала производится в той же спектральной области, в которой будет осуществляться кодирование моносигнала. В рассматриваемом случае банком фильтров, используемым для кодирования сигнала, является модифицированное дискретное косинусное преобразование. Для анализа потребуется также мнимая составляющая данного преобразования, которая вычисляется путем применения модифицированного дискретного синусного преобразования к сигналу. Обе части данного преобразования образуют комплексный спектр, однако для того чтобы восстановить временной сигнал, достаточно лишь его вещественной части.

Выберем один сигнал из стереопары, например левый, в качестве опорного и найдем параметры для формирования (предсказания) другого канала, правого. В качестве параметров используем пару (g, φ) , при этом будем оценивать только ошибку предсказания вещественной части сигнала. Имеем задачу

$$E(g, \varphi) = \sum_{k=1}^q \left[a_k^R \cos \varphi_k^R - g a_k^L \cos(\varphi_k^L + \varphi) \right]^2 \rightarrow \min_{g, \varphi}.$$

Вычислим частные производные

$$\frac{\partial E(g, \varphi)}{\partial g} = \sum_{k=1}^q \left[a_k^R \cos \varphi_k^R - g a_k^L \cos(\varphi_k^L + \varphi) \right] a_k^L \cos(\varphi_k^L + \varphi),$$

$$\frac{\partial E(g, \varphi)}{\partial \varphi} = \sum_{k=1}^q \left[a_k^R \cos \varphi_k^R - g a_k^L \cos(\varphi_k^L + \varphi) \right] a_k^L \sin(\varphi_k^L + \varphi).$$

Приравняв частные производные к нулю, получим систему уравнений

$$\left. \begin{aligned} g &= \frac{\sum_{k=1}^q a_k^R a_k^L \cos(\varphi_k^L + \varphi) \cos \varphi_k^R}{\sum_{k=1}^q (a_k^L)^2 \cos^2(\varphi_k^L + \varphi)}, \\ g &= \frac{\sum_{k=1}^q a_k^R a_k^L \sin(\varphi_k^L + \varphi) \cos \varphi_k^R}{\sum_{k=1}^q (a_k^L)^2 \cos(\varphi_k^L + \varphi) \sin(\varphi_k^L + \varphi)}. \end{aligned} \right\} \quad (1)$$

Данная система уравнений не может быть решена аналитически. Однако так как система имеет вид

$$\left. \begin{aligned} g &= f_1(\varphi); \\ g &= f_2(\varphi), \end{aligned} \right\}$$

то значение угла φ , являющееся решением экстремальной задачи

$$|f_1(\varphi) - f_2(\varphi)| \rightarrow \min_{\varphi}, \quad (2)$$

одновременно является решением системы (1).

Задачу (2) предлагается решать методом перебора. Для этого выберем параметр $\nu \in \mathbb{N}$, определяющий точность решения системы уравнений (1), и множество $\Omega_{\nu} = [0 : 2\pi/\nu : 2\pi]$. Путем перебора всех $\varphi \in \Omega_{\nu}$ решим задачу (2). Значение коэффициента g вычисляется из первого уравнения системы (1).

Полученные стереопараметры (g, φ) квантуются и передаются в поток. Обозначим деквантованную пару через $(\hat{g}, \hat{\varphi})$. Разностный сигнал вычисляется как

$$D_k = a_k^R \cos \varphi_k^R - g a_k^L \cos(\varphi_k^L + \varphi),$$

его вещественная часть обрабатывается монокодером. Как правило, разностный сигнал имеет небольшую энергию, следовательно, не требует значительных затрат на кодирование.

Заключение. Представлены два новых метода кодирования стереофонического аудиосигнала. Каждый из методов имеет свою область применения. Первый может быть применен, когда требуется передать стереофоническую картину при наименьших затратах на кодирование, второй — когда необходимо получить идеальное качество восстановленного стереосигнала.

Описание методов квантования и эффективного кодирования параметров стереосигнала выходит за рамки данной статьи. Однако заметим, что при использовании в стандартном аудиокодере представленных методов наблюдается улучшение перцептуального качества восстановленного аудиосигнала по сравнению с известными аналогами для большинства тестовых последовательностей.

СПИСОК ЛИТЕРАТУРЫ

1. ISO/IEC 13818-7. Information Technology — Generic Coding of Moving Pictures and Associated Audio Information — Part 7: Advanced Audio Coding (AAC). 1997.
2. Herre J., Brandenburg K., Lederer D. Intensity stereo coding // Proc. of the 96th AES Convention, Amsterdam, 1994. (Preprint N 3799).
3. Faller C. Parametric coding of spatial audio // Proc. of the 7th Intern. Conf. on Digital Audio Effects (DAFx-04), Naples, Italy, Oct. 2004.
4. Johnston J. D. A filter family designed for use in quadrature mirror filter banks // Proc. IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing. 1980. April. P. 291—294.
5. Temerinac M., Edler B. Overlapping block transform: window design, fast algorithm, and an image coding experiment // IEEE Trans. on Communications. 1995. Vol. 43, N 9. P. 2417—2425.

Сведения об авторах

- Кирилл Валерьевич Юрков** — канд. техн. наук, доцент; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра информационных систем; E-mail: yourkovkirill@mail.ru
- Сергей Евгеньевич Петров** — Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, НИИ наукоемких компьютерных технологий; мл. науч. сотрудник; E-mail: petrovse@mail.ru

Рекомендована кафедрой
информационных систем

Поступила в редакцию
08.07.10 г.