ИНФОРМАЦИОННЫЕ КРИТЕРИИ ЭФФЕКТИВНОСТИ БАНКОВ ФИЛЬТРОВ

Кудряшов Б.Д., Осипов К.С.

Санкт-Петербургский государственный университет аэрокосмического приборостроения

Рассматривается задача рационального выбора длины преобразования при кодировании аудио сигнала. Вместо обычного критерия энергетической эффективности предлагается использовать более сложный информационный критерий, более адекватно отображающий ориентировочные битовые затраты при заданном искажении. Критерий основан на аппроксимации одномерного распределения при помощи модели обобщенного гауссовского распределения.

1. Введение

В настоящее время основные проблемы в области сжатия аудио информации связаны с построением банков фильтров для преобразования временного сигнала в более пригодный для обработки и сжатия частотный вид, поисками хорошей психоакустической модели (ПАМ) [1] для определения того, какая часть сигнала должна быть передана более точно, а чем можно пренебречь, а также с определением эффективного метода квантования передаваемых данных. В данной работе рассматривается проблема выбора хорошего банка фильтров и методики его адаптации к изменяющимся свойствам аудио сигнала.

Банк фильтров служит для преобразования сигнала из временной формы в частотную. Обычно он представляется в виде множества полосовых фильтров, покрывающих весь доступный частотный диапазон. Он делит всю спектральную область на частотные поддиапазоны и генерирует серии коэффициентов (кадры) в последовательном временном порядке.

Звуковые сигналы представляет собой упругие гармонические колебания, возникающие в воздушной среде, поэтому соответствующие им электрические сигналы хорошо описываются раложением в базисе, построенном на использовании гармонических функций.

Одним из наиболее часто используемых для кодирования банков фильтров является модифицирование дискретное косинусное преобразование (МДКП) с перекрытием 1/2. В матричном виде преобразование может быть записано как

$$\mathbf{f} = \mathbf{x} \times \mathbf{W} \times \mathbf{T} \times \mathrm{DCTIV}, \qquad (1)$$

где $\mathbf{x}$ – вектор входных данных, $\mathbf{W}$ – диагональная матрица окна (обычно синусоидального), $\mathbf{T}$ – матрица предобработки и DCTIV – матрица преобразования DCTIV.

В общем виде длина входного вектора $\mathbf{x}$ равна сумме длин двух полукадров (степени двойки) $2N_p$ и $2N_c$. Матрица окна может быть записана как

$$\mathbf{W} = \begin{Vmatrix} \mathbf{W_p} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_c} \end{Vmatrix}, \qquad (2)$$

где $\mathbf{W}_p$ и $\mathbf{W}_c$ – квадратные диагональные матрицы, диагональные элементы $\mathbf{W}_p$ соответствуют первой половине оконной функции N_p, и диагональные элементы $\mathbf{W}_c$ соответствуют второй половине оконной функции длины N_c, $\mathbf{0}$ – матрица из всех нулей соответствующего размера. Таким образом, длина преобразования DCTIV равна $N_p + N_c$. Видно, что в этом преобразовании число получившихся коэффициентов будет равно половине длины входного вектора. В совокупности с половинным перекрытием можно утверждать что число частотных коэффициентов равно числу входных сэмплов.

Одна из важных проблем при выборе параметров банка фильтров N_p и N_c связана с тем, что в последовательности звуков каждый из них обычно обладает специфическими свойствами, поскольку звуки могут порождаться различными источниками. Учет этих различий уже на этапе преобразования может быть полезен для получения более компактного представления. Таким образом, можно сделать вывод, что эффективный аудио кодер должен учитывать изменения свойств сигнала во времени.

Для решения проблемы нестационарности можно разбивать сигнал на участки с различными свойствами, например путём управления длиной преобразования (длиной кадра). Управление переключением длин возложено на детектор атаки, он анализирует какие-либо свойства сигнала и по некоторому эмпирическому алгоритму расставляет границы кадров.

Понятно, что требования к сложности алгоритмов кодирования накладывают ограничения на выбор длины кадра – она не может быть абсолютно любой. Чаще всего выбор делается между всего двумя типами – длинным и коротким кадром, в более сложном случае – несколькими фиксированными длинами.

2. Модель

Выбор обобщенного гауссовского распределения вероятностей обусловлен тем, что экспериментальные исследования убеждают в применимости распределений этого класса для описания моделей случайных величин в задачах кодирования аудио информации. Согласно этой модели, источник порождает случайную величину, плотность распределения вероятностей которой имеет вид

$$f(x) = \frac{\alpha\eta(\alpha,\sigma)}{2\Gamma(1/\alpha)}\exp\left\{-\left(\eta(\alpha,\sigma)|x-m|\right)^{\alpha}\right\}, \quad (3)$$

где m,σ – математическое ожидание и среднеквадратическое отклонение, α – параметр распределения, $\Gamma(\cdot)$ – гамма функция, а значения функции $\eta(\alpha,\sigma)$ вычисляются по формуле

$$\eta(\alpha,\sigma) = \sigma^{-1}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right]^{1/2}. \quad (4)$$

Параметр α характеризует экспоненциальную скорость убывания хвостов плотности, именно этот параметр меняется в зависимости от свойств сигнала. Отметим, что значениям $\alpha = 1,2$ соответствуют распределение Лапласа и нормальное распределение. Для решения большинства задач, возникающих при кодировании звука, можно использовать $\alpha = 0.25$.

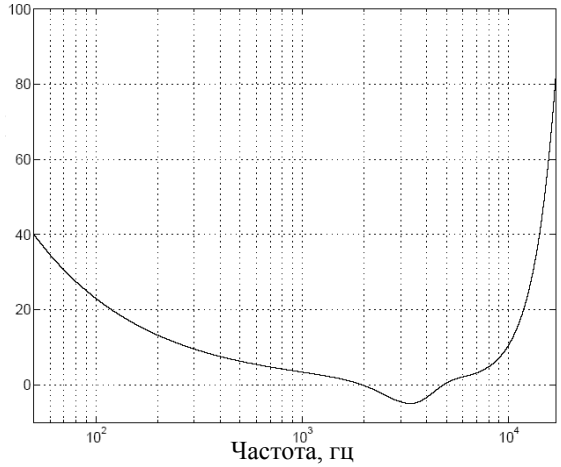


Рис.1 Кривая слышимости

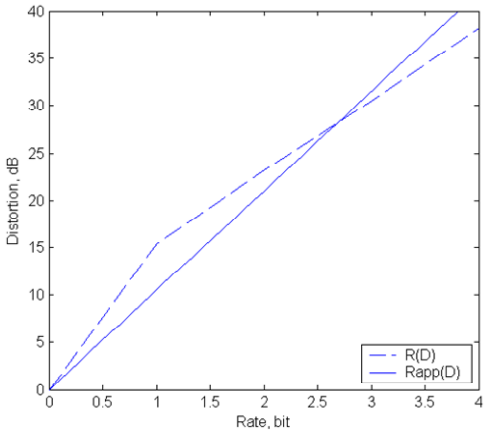


Рис. 2. Кривая R(D) и её аппроксимация

Из теории построения «психоакустической модели» для аудио кодеров известно важное свойство, определяющее чувствительность человеческого уха в различных частотных диапазонах. Приведённая на рис. 1 абсолютная кривая слышимости определяет уровень минимальный слышимого чистого тона в зависимости от его частоты.

В психоакустических моделях распределения бит абсолютный порог слышимости часто используют как меру, дающую неслышимое на слух искажение. В этом случае SPL-нормализованный спектр разбивается на 25 диапазонов, заданных барковским законом. В каждом диапазоне находится минимальное значении порога слышимости, и максимальная амплитуда спектра в дБ. Разность между этими значениями даёт отношение сигнал-шум в дБ в диапазоне, при котором искажение будет неслышным.

3. Скорость кодирования

Для случайного ансамбля X предельная скорость кодирования при заданной ошибке D определяется ε-энтропией источника при $\varepsilon = D$, ее обозначают как $H_{\varepsilon}(X)$. Эта же величина часто называется функцией скорость-искажение $R(D)$ [2].

График этой функции для используемого обобщённого гауссовского процесса с параметром $\alpha = 0.25$ может быть получен с использованием функции Блейхута [3].

Для упрощения расчётов аппроксимируем полученную кривую прямой $R_{app}(D)$ с параметром

$$R_{app}(D) = 0.095\log(D), \text{ бит.} \quad (5)$$

Графики кривой и аппроксимирующей прямой приведены на рис. 2.

4. Информационный критерий эффективности преобразования

Для построения информационного критерия используем описанные выше методы построения психоакустических моделей, а также зависимости $R(D)$ полученные в п.3.

Модель ПАМ даёт для каждого частотного поддиапазона отношение сигнал-шум, при котором искажение считается неслышимым. Затем при помощи аппроксимированной зависимости $R_{app}(D)$ мы можем оценить скорость кодирования для данного диапазона. Битовые затраты для всего кадра являются суммой оценки по всем диапазонам.

$$Bits = \sum_{i=0}^{25} R_{app}(D_i) L_i \text{, бит,} \qquad (6)$$

где D_i – допустимый уровень искажения в i-м диапазоне, L_i – его длина.

В качестве примера возьмём 3 банка фильтров, построенных на преобразовании МДКП: с длиной кадра 1024 частотных коэффициентов; с длиной 128 коэффициентов; третий банк имеет две длины: 2048 и 128 коэффициентов, и возможность переключения между ними по некоторому эмпирическому алгоритму.

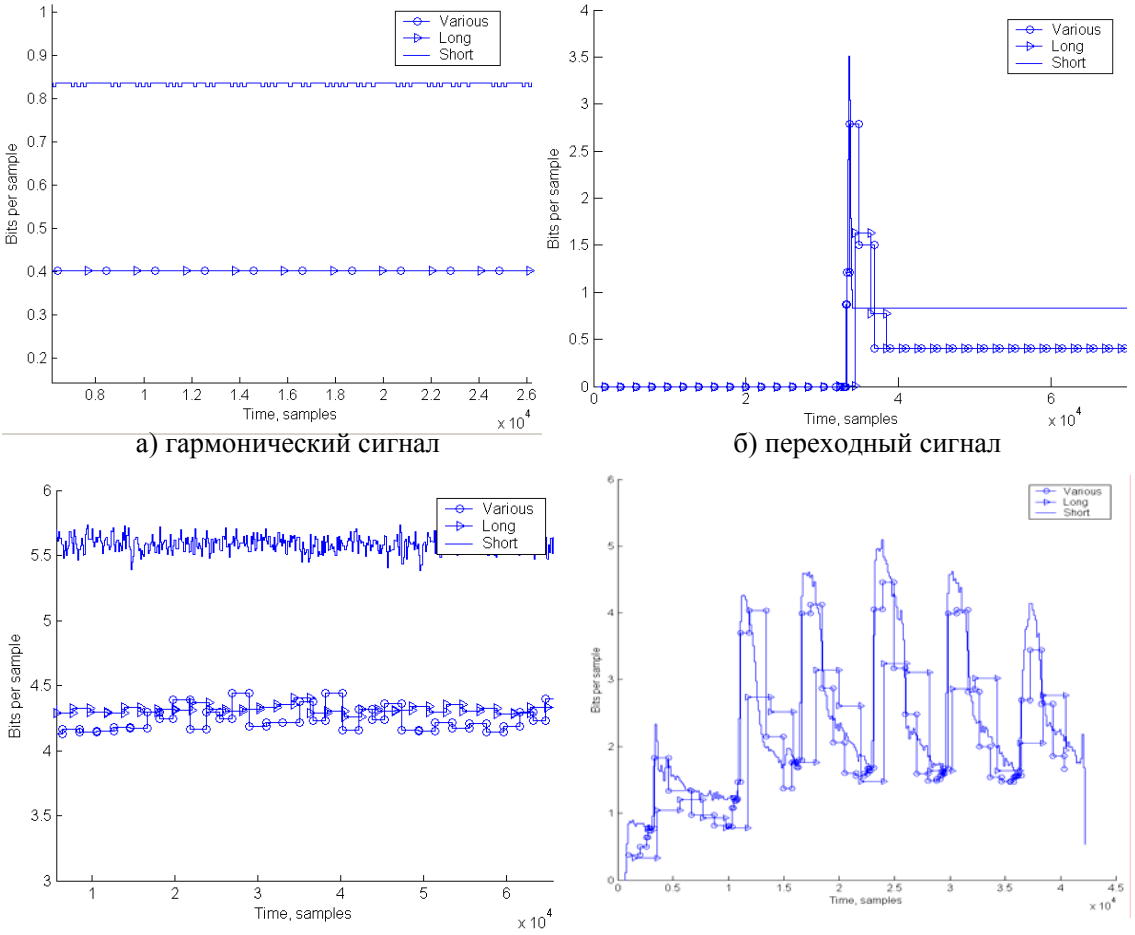
Рассмотрим несколько типичных видов звуковых сигналов, и проанализируем необходимое для кодирования количество бит по описанному выше критерию. На графиках приведено число бит на отсчёт, необходимое для передачи сигнала без слышимых искажений.

Гармонический синусоидальный сигнал с частотой 3кГц, рис 3, а. Банка фильтров с переменной длиной кадра эквивалентен банку с длинными кадрами и выигрывает над только короткими 0.41 бита на отсчёт.

Переходный сигнал, рис 3. б. Участок тишины, за которым следует сигнал, имитирующий звуковую атаку. Средний выигрыш банка фильтров с переменной длиной над только длинными кадрами составляет 0.013 бита на отсчёт, над только короткими 0.3 бита на отсчёт.

Гауссовский шум, рис 3. в. Средний выигрыш банка фильтров с переменной длиной над только длинными кадрами составляет 0.184 бита на отсчёт, над короткими 1.48 бита на отсчёт.

Реальный звуковой сигнал, рис 3. г., поп-музыка. Средний выигрыш банка фильтров с переменной длиной над только длинными кадрами составляет 0.15 бита на отсчёт, над только короткими 0.43 бита на отсчёт.



а) гармонический сигнал

б) переходный сигнал

в) шумовой сигнал

г) реальный сигнал

Рис. 3. Зависимости распределения бит для выбранных фильтров.

5. Заключение

Приведенные в работе результаты анализа и моделирования позволяют сформулировать подход к обоснованному выбору длины преобразования на основании анализа обработки фрагмента аудио сигнала. Предложенный информационный критерий дает объективную оценку потенциальных битовых затрат для каждого конкретного банка фильтров. Определенным недостатком данного подхода является повышенная сложность реализации.

С другой стороны, как показали приведенные выше вычисления, за счет применения данного критерия возможно получение выигрыша при кодировании реального сигнала до 0.15 бит на отсчёт, что вполне компенсирует увеличение вычислительной сложности.

Помимо этого, данный критерий можно использовать в комбинации с методами обнаружения переходных сигналов во временной области, когда более простой эмпирический алгоритм выдаёт несколько правдоподобных гипотез, и окончательный выбор осуществляется при помощи предложенного критерия.

Литература

1. Painter T., Spanias A. Perceptual Coding of Digital Audio. Proceedings of the IEEE, v. 88, N 4, 2000, pp.. 451-513

2. Колесник В. Д., Полтырев Г. Ш. Курс теории информации: Учебное пособие для вузов. М.: Наука, 1982. 416 с.

3. Blahut R. E., “Computation of Channel Capacity and Rate-Distortion Functions”, IEEE Trans. Inform. Theory, 18, No 4, pp. 460-473, Jul., 1972.

INFORMATION MEASURE OF FILTERBANK EFFICIENCY

Kudryashov B., Osipov K

State university of aerospace instrumentation, Saint-Petersburg

The problem of optimum selection of frame length for transform coding of audio signals is considered. Conventional criteria typically are based on the measurements of the signal energy in time domain (or in some subbands of audio signal) and detection of abrupt changes. Unlike these approaches more sophisticated criterion is suggested. This criterion takes into account the estimates on the number of bits required for achieving required approximation precision.

The proposed information measure is based on stochastic modeling of the transformed audio signal. The generalized Gaussian distribution with varying parameters is considered as a probability density function. An estimate on the bit rate is obtained using rate-distortion function for a given distribution with approximation error value selected taking into account the audibility of distortions. In other words, the required approximation precision is computed using psychoacoustic model, which takes into account sensitivity of the human ear with respect to different kinds of distortion depending on the magnitude and frequency range.

The rate-distortion function $R(D)$ (or ε-entropy, $\varepsilon = D$) of the analog source determines the theoretical lower limit of the achievable bit rate under constraint D on the average distortion. The rate-distortion function for given probability distribution can be computed using Blahut algorithm.

To illustrate the behavior of the information measure a set of examples is presented. Among them are: pure tone signal, imitation of sound attack, noise signal, and real audio waveform. The bit expenses for different strategies of frame switching are computed.

The presented results lead to following conclusions:

- The proposed information measure provides reliable estimates of the bit expenses for different frame length switching strategies and therefore this measure can be used as a criterion for selecting proper frame length when the input signal parameters are changed or sound attack is detected;
- For signal with sound attacks switching to short frame provides substantial saving of required number of bits, and for stationary signal long frame length should be preferred;
- Exploiting considered information measure as a frame switching criteria can lead to high computational complexity since a few trials of signal transformation could be required for each frame. Nevertheless, proposed approach seems to be promising because simulation results demonstrated reduction of bit rate about 0.15 bits per sample of audio signal.