

Федеральное агентство связи

**Федеральное государственное образовательное бюджетное учреждение
высшего профессионального образования**

**ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ**

**ЭЛЕКТРОННАЯ
БИБЛИОТЕЧНАЯ СИСТЕМА**

Самара

На правах рукописи

СТЕФАНОВ Михаил Александрович

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ СЖАТИЯ ЦИФРОВОЙ
АУДИОИНФОРМАЦИИ С УЧЕТОМ СВОЙСТВ СЛУХОВОГО
АНАЛИЗАТОРА ЧЕЛОВЕКА**

Специальность 05.12.13 –
Системы, сети и устройства телекоммуникаций

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Самара – 2009

Работа выполнена в Государственном образовательном учреждении высшего профессионального образования «Поволжский государственный университет телекоммуникаций и информатики» (ГОУВПО ПГУТИ)

Научный руководитель: доктор технических наук,
профессор Акчурин Э.А.

Официальные оппоненты: доктор технических наук,
доцент Горячкин О.В.

кандидат технических наук,
доцент Засов В.А.

Ведущая организация: ГОУВПО Самарский государственный технический университет (СГТУ)

Защита состоится «28» декабря 2009 г. в 13 часов на заседании диссертационного совета Д219.003.02 при Поволжском государственном университете телекоммуникаций и информатики по адресу: 443010, г. Самара, ул. Льва Толстого, 23.

С диссертацией можно ознакомиться в библиотеке ГОУВПО ПГУТИ.

Автореферат разослан «26» ноября 2009 г.

Ученый секретарь
диссертационного совета Д219.003.02
доктор технических наук, доцент

Мишин Д.В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Современный этап развития систем цифровой обработки и передачи аудиоинформации характеризуется как увеличением потока передаваемой информации, так и повышением требований к ее качеству. В связи с этим одной из важнейших задач является повышение эффективности компрессии цифровых аудиоданных при условии высокого качества субъективного восприятия сжатого звука.

К настоящему времени в стандартах *MPEG (Moving Pictures Expert Group)* и *ATSC (Advanced Television System Committee) Dolby AC-3 (A/52)* рекомендуются алгоритмы высококачественного сжатия звуковых сигналов (ЗС), обеспечивающие цифровые скорости от 64 кбит/с для мультимедиа приложений до 384 кбит/с для цифрового стереофонического вещания. Нестандартные алгоритмы в форматах *PAC, VQF, WMA, Ogg Vorbis* и другие по утверждениям их авторов позволяют без существенного снижения качества субъективного восприятия сжатого звука довести цифровую скорость до 96 кбит/с. Таких успехов удалось добиться за счет обработки цифровых аудиоданных в частотной области, для чего используются дискретные ортогональные преобразования (ДОП). В свою очередь, такое радикальное изменение принципов обработки ЗС стало возможным благодаря результатам экспериментальных исследований психоакустических свойств слуха. Наиболее полно эти исследования провели *E. Zwicker, R. Feldtkeller*. Вместе с тем прикладной характер эти результаты получили благодаря усилиям зарубежных ученых *K. Brandenburg, G. Stoll, G. Theile* и др. В России следует выделить работы Ю.А. Ковалгина и А.М. Синильникова.

Однако дальнейшее развитие высококачественных систем компрессии цифровых аудиоданных сдерживается отсутствием условий эффективного использования психоакустических свойств слуха. Так, в стандартах *MPEG* и *DOLBY AC-3 (A/52)* эффект маскировки учитывается только для оценки допустимой степени округления. А в известных предложениях по повышению эффективности сжатия не учитываются частотные характеристики ДОП и особенности восприятия слухом сжатого спектра. В результате не представляется возможным предметно говорить о более эффективных алгоритмах кодирования спектра ЗС.

В связи с этим актуальной является задача выработки обоснованных рекомендаций относительно ДОП, перспективных для целей сжатия ЗС и условий повышения производительности высококачественных систем компрессии цифровых аудиоданных.

Цель и задачи работы. Целью диссертации является повышение эффективности сжатия цифровой аудиоинформации. Для достижения указанной цели в диссертационной работе поставлены следующие задачи:

- сформулировать обоснованные требования и выработать рекомендации относительно ДОП, перспективных для эффективной и высококачественной компрессии цифровых аудиоданных;
- исследовать ДСО коэффициентов перспективных ДОП с учетом их частотных свойств и основных психоакустических свойств слуха;

- исследовать эффективность прореживания коэффициентов перспективных ДОП;
- разработать методики и провести экспериментальное исследование эффективности сжатия цифровых аудиоданных с прореживанием спектра.

Методы исследования. При решении поставленных задач использовались методы спектрального анализа, теории вероятностей и математической статистики, теории цифровой обработки сигналов, результаты экспериментальных исследований психоакустических свойств слуха и статистических свойств ЗС, компьютерное моделирование.

Научная новизна работы заключается в следующем:

- разработан метод оценки распределения допустимой степени округления коэффициентов гармонических дискретных ортогональных преобразований, учитывающий их частотные характеристики и свойства слухового анализатора человека;
- получены аналитические выражения для оценки вероятности маскировки коэффициентов гармонических дискретных ортогональных преобразований любым участком кривой маскировки;
- разработан алгоритм прореживания коэффициентов гармонических дискретных ортогональных преобразований, учитывающий особенности восприятия слухом компонент сжатого спектра;
- разработан метод оценки частотной области целесообразной для прореживания в заданном рабочем диапазоне частот.

Личный вклад. Основные научные положения, теоретические выводы и рекомендации, а также методики экспериментальной обработки цифровых аудиоданных, содержащиеся в диссертационной работе, получены автором самостоятельно.

Практическая ценность и реализация результатов работы. Полученные аналитические выражения и зависимости способствуют повышению производительности систем сжатия цифровых аудиоданных при сохранении высоких качественных показателей.

Результаты работы могут быть использованы при разработке новых и усовершенствовании существующих цифровых технологий передачи и обработки цифровой аудиоинформации. Разработанные методики экспериментальной обработки цифровых аудиоданных могут быть использованы для оценки эффективности новых алгоритмов сжатия.

Основные теоретические и практические результаты, полученные в работе, использованы ФГУП СОНИИР и внедрены в учебный процесс ГОУВПО ПГУТИ, что подтверждено соответствующими актами.

Апробация работы. Основные результаты и положения работы обсуждались на VIV – XVI Российских научных конференциях (Самара, 2007 – 2009), 6 Международной научно-технической конференции «Проблемы техники и технологий телекоммуникаций» (Уфа, 2005), VII Международной научно-технической конференции «Проблемы техники и технологий телекоммуникаций» (Самара, 2006), IX Международной научно-технической конференции «Проблемы техни-

ки и технологий телеком-2008), V Международной конференции «Методы и сред-логическими процессами»

Публикации. По теме но 11 печатных работ, в том изданиях, входящих в пере-

текстов докладов на Российских и Международных конференциях.

Основные положения, выносимые на защиту:

- метод оценки распределения допустимой степени округления коэффициентов гармонических дискретных ортогональных преобразований, учитывающий их частотные характеристики и свойства слухового анализатора человека;
- аналитические выражения для оценки вероятности маскировки коэффициентов гармонических дискретных ортогональных преобразований любым участком кривой маскировки;
- метод оценки частотной области целесообразной для прореживания, в заданном рабочем диапазоне частот;
- результаты компьютерного моделирования обработки цифровой аудиоинформации.

Структура и объем работы. Диссертационная работа состоит из введения, четырех глав, заключения, списка литературы и приложений. Работа содержит 146 страниц машинописного текста, 64 рисунка, 10 таблиц. Список литературы включает 168 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы исследования, сформулированы цели и задачи работы, перечислены основные научные результаты диссертации, положения, выносимые на защиту, определены практическая ценность и области применения результатов, приведены сведения об апробации работы, структуре и объеме диссертации.

В первой главе на основе психофизиологических свойств слуха и характерных особенностях работы систем обработки и передачи ЗС формулируются требования, которым должны удовлетворять ДОП при использовании их в целях высококачественного сжатия цифровой аудиоинформации и проводится сравнительный анализ наиболее известных ДОП.

Распределение амплитуд колебаний вдоль основной мембраны органа слуха экспериментально исследовано Бекеша (рис. 1). Оказалось, что тон возбуждает не одно поперечное волокно, настроенное на данную частоту, а целый участок мембраны. Тем не менее, в итоге ощущается только сам испытательный тон. Эти результаты позволяют сделать вывод, что в высококачественных системах сжатия ДОП должны допускать интерпретацию своих коэффициентов как тональных компонент спектра, так и полосного шума, интенсивность которого сконцентрирована на частоте соответствующего коэффициента. Обе интерпретации отнюдь не противоречивы, поскольку с точки зрения электроакустики распределение интенсивности полосного шума безразлично. Она может быть распределена равномерно или сосредоточена в одной гармонике.

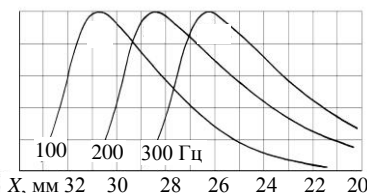


Рис. 1. Поперечные колебания основной мембраны

муникаций» (Казань, научно-технической кон-ства управления техно-(Саранск, 2009).

диссертации опубликована-числе 2 статьи в научных-чень ВАК, и 9 тезисов и

Согласно результатам экспериментальных исследований эффекта маскировки (подавление тонов меньшей интенсивности тоном большей интенсивности) слухового анализатора, колебания поперечных волокон основной мембраны некоррелированы. Следовательно, этим свойством должны обладать и коэффициенты ДОП. Кроме того, кривая порога слышимости при маскировке, иначе кривая маскировки (КМ), при определенных условиях может распространяться почти на весь слышимый диапазон частот (рис. 2). Тем самым обуславливается необходимость глобальной декорреляции коэффициентов ДОП.

Эти требования следует дополнить еще одним, учитывающим работу подавляющего числа систем обработки и передачи цифровой аудиоинформации в режиме реального времени – ДОП должно обладать быстрым вычислительным алгоритмом.

Наибольшую эффективность сжатия цифровых аудиоданных можно ожидать при использовании преобразования Карунена-Лоэва, коэффициенты которого полностью декоррелированы. Более того, по критерию среднеквадратичной ошибки оно является оптимальным для представления стационарных процессов, Карунена-Лоэва производится по интегральному уравнению корреляционной функции применительно к ЗС на инвала стационарности, для которого придется строить свой базис функции. Кроме того, разложение Карунена-Лоэва не имеет быстрых вычислительных алгоритмов.

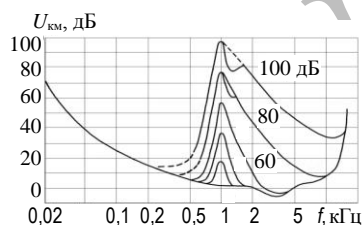


Рис. 2. Кривые маскировки

мальным для представления когда разложение Карунена-Лоэва производится по интегральному уравнению корреляционной функции применительно к ЗС на инвала стационарности, для которого придется строить свой базис функции. Кроме того, разложение Карунена-Лоэва не имеет быстрых вычислительных алгоритмов.

Преобразование Хаара обладает быстрыми вычислительными алгоритмами. Однако, как известно, свойством глобальной декорреляции обладают только небольшое число первых коэффициентов.

Преобразование Уолша-Адамара также имеет быстрые вычислительные алгоритмы. Упорядоченное по Адамару, оно обеспечивает глобальную декорреляцию, но каждая точка его энергетического спектра представляет собой энергетическое содержание не одной, а группы частот. В результате восприятие на слух сжатого и исходного спектров будет резко отличаться. При упорядочивании по Уолшу обеспечивается и глобальная декорреляция, и адекватный ЗС энергетический спектр. Вместе с тем функции Уолша, положенные в основу преобразования, представляют собой прямоугольные, причем как периодические, так и непериодические волны. Столь резкие изменения частоты и амплитуды сигнала ощущаются на слух как щелчки.

Вейвлет-преобразование дает хорошее разрешение по частоте и плохое по времени в низкочастотной области. Обратное в высокочастотной области. Эта особенность хорошо согласуется с результатами экспериментальных исследований спектральной плотности средней мощности для ЗС различных жанров. Однако, во-первых, точность анализа посредством вейвлет-преобразования определяется степенью соответствия форм вейвлета и сигнала. Форму же ЗС на малых временных интервалах предсказать невозможно. Во-вторых, вейвлет-

преобразование дает оценку суммарной интенсивности, сосредоточенной в некоторой полосе частот анализируемого сигнала. А, как известно, слуховой анализатор сигнал на полосы не разделяет.

Из класса гармонических преобразований наиболее известными являются преобразования Фурье, Хартли и косинусное преобразование (ДКП). Первые два преобразования по свойствам абсолютно идентичны. Разница лишь в том, что второе не требует комплексной арифметики. Поэтому в дальнейшем они обозначаются как ДПФ-Х.

В целом же все эти преобразования обладают быстрыми вычислительными алгоритмами и обеспечивают глобальную декорреляцию своих коэффициентов. На рис. 3 приведены их нормированные АЧХ $|H_k(i\omega)|_н = |H_k(i\omega)|/|H_k(i\omega_k)|$. Видно, что коэффициенты $K_x(k)$ этих преобразований можно интерпретировать как полосный шум, интенсивность которого сосредоточена на частоте f_k . Вместе с тем в силу слабого затухания амплитуды боковых лепестков и равновероятности знаков соответствующих спектральных компонент исходного сигнала в первом приближении $K_x(k)$ можно считать некоррелированными. Однако для ДКП это справедливо только для четных и нечетных коэффициентов в отдельности.

Таким образом, перспектива цифровой аудиоинформационно-гармонические ДОП. При этом коэффициенты ДКП как то-спектра их целесообразно раз-признаку четности.

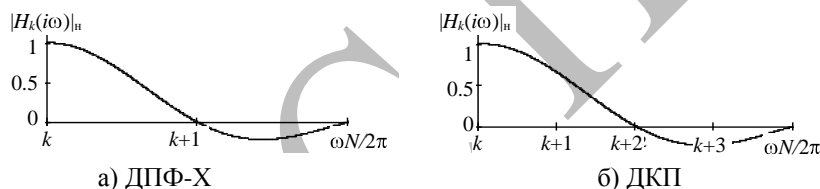


Рис. 3. АЧХ коэффициентов гармонических ДОП

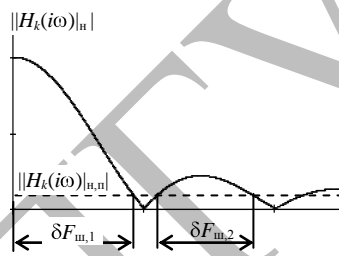


Рис. 4. Составляющие $\Delta F_{ш,k}$

тивными для целей сжатия следует признать в случае интерпретации отдельных компонент делить на два потока по

Во второй главе с учетом частотных свойств ДОП и основных психоакустических свойств слуха разрабатывается методика оценки распределения ДСО ко-

эффициентов ДОП в пределах спектральной выборки.

Естественным критерием ограничения разрядности $q1_k$ модуля $K_x(k)$ является коэффициент минимально ощущаемой на слух прямоугольной амплитудной модуляции полосного шума m_k , концентрируемого на частоте f_k . В свою очередь, m_k является функцией двух параметров – ширины полосы шума $\Delta F_{ш,k}$ и частоты его модуляции $f_{мод} = 1/T_{п}$, где $T_{п}$ – длительность интервала преобразования. Известное условие оценки $T_{п}$ получено при участии автора данной работы и позволяет оценить оптимальное значение соответствующего параметра m_k : $f_{мод} = 130$ Гц для музыкальных сигналов и 100 Гц для речевых сигналов.

Второй параметр m_k обычно принимается равным рабочей полосе ЗС. Однако (рис. 3) степень влияния спектральных компонент на величину $K_x(k)$ различна. Следовательно, должно существовать пороговое значение $|H_k(i\omega)|_{н,п}$ (рис. 4), разделяющее значимые в этом смысле компоненты спектра исходного сигнала s_m

($s_m \in \delta F_{ш,j}$) от незначимых s_n ($s_n \notin \delta F_{ш,j}$), где $\delta F_{ш,j}$ – часть полосного шума, обусловленная j -м лепестком АЧХ. Тогда и поскольку фрагментарный характер шума для m_k значения не имеет

$$\Delta F_{ш,k} = \sum_j \delta F_j \cdot \quad (1)$$

Учитывая, что занижение $q1_k$ ведет к изменению взаимодействия компонент нового (после обратного преобразования) спектра, $|H_k(i\omega)|_{н,п}$ целесообразно искать на основе эффекта маскировки.

Пусть для $K_x(k)$ известна вероятность маскировки G_k . В плоскости слышимости ей можно сопоставить область безусловной маскировки, ограниченной сверху средним порогом слышимости при маскировке $U_{пс,k}$. Тогда вероятность G_k можно переопределить как $P\{U_k < U_{пс,k}\} = dec(U_{пс,k}/20)/dec(D_k/20)$, где U_k – уровень $K_x(k)$ в плоскости слышимости, ограниченной сверху динамическим диапазоном изменения сигнала, а $dec(x) = 10^x$. Отсюда $dec(U_{пс,k}/20) = P\{U_k < U_{пс,k}\} \cdot dec(D_k/20)$. В наихудшем случае s_n обуславливает приращение $u_k = dec(U_k/20)$, не превышающее по абсолютной величине $\|H_k(i\omega_n)\|_{н} \cdot dec(D_n/20)$. При этом s_n будет незначимой для $K_x(k)$, если $\|H_k(i\omega_n)\|_{н} \cdot dec(D_n/20) < dec(U_{пс,k}/20)$. Таким образом, условие оценки ширины полос $\delta F_{ш,j}$, составляющих $\Delta F_{ш,k}$, запишется в виде:

$$\|H_k(i\omega_m)\|_{н} \geq G_k \cdot dec(D_k/20) / dec(D_m/20) = \|H_k(i\omega)\|_{н,п}. \quad (2)$$

С целью оценки G_k получена более адекватная по форме и вместе с тем достаточно точная (коэффициент детерминации $R^2 \rightarrow 1$) аппроксимация КМ:

$$\left. \begin{aligned} U_{км,i}^{(n)} &= -54|fo_{k,i}|^{1,6} + U_i + a_{m,i}, \quad U_{км,i}^{(2)} = -3fo_{k,i} + (0,05fo_{k,i} + 0,76)U_i + a_{m,i}, \\ U_{км,i}^{(1)} &= -18|fo_{k,i}|^{1,6} + U_i + a_{m,i}, \quad U_{км,i}^{(3)} = -\alpha_1 \lg[1 + \alpha_2 fo_{k,i} \exp(-\alpha_3 U_i)] + U_i + a_{m,i}, \end{aligned} \right\} \quad (3)$$

где верхний индекс означает участок КМ $K_x(i)$ уровнем U_i , относительная частота $fo_{k,i} = (f_k - f_i)/\Delta F_i$ и ΔF_i – ширина частотной группы (области волосковых клеток, возбуждаемых минимально слышимым $K_x(i)$), $a_{m,i}$ – коэффициент маскировки (определяет уровень пика КМ относительно уровня маскирующего тона или шума), а коэффициенты $\alpha_1 - \alpha_3$ определяются эмпирически. Кроме того, определяется безусловная вероятность $P\{U_k < U_{км}^{(x)}(fo_{k,i}, U_i)\}$ маскировки $K_x(k)$ участком x КМ $_i$, для чего соответствующая область ее значений на частоте f_k приводится к единичному квадрату. В результате при маскировке левой ветвью и первыми двумя участками правой ветви эта вероятность равна 0,5, а третьим участком

$$P\{U_k < U_{км}^{(3)}(fo_{k,i}, U_i)\} = \frac{\ln[\alpha_2 + \exp(\alpha_3 \cdot U_{i,гп1})]}{\beta} + \frac{1}{\beta \cdot \alpha_3 \cdot U_{i,гп2} - U_{i,гп1}} \times$$

$$\left\{ \begin{array}{l}
-0.5 \cdot \alpha_3 \cdot U_{i, \text{rp}2} - U_{i, \text{rp}1} \cdot [2 \cdot \ln(\alpha_2) - \alpha_3 \cdot U_{i, \text{rp}2} - U_{i, \text{rp}1}] + \\
\sum_{n=1}^{\infty} \frac{-1^n \cdot \exp(\alpha_3 U_{i, \text{rp}2} n)}{n^2 \cdot \alpha_2^n} - \sum_{n=1}^{\infty} \frac{-1^n \cdot \exp(\alpha_3 U_{i, \text{rp}1} n)}{n^2 \cdot \alpha_2^n}, U_{i, \text{rp}2} < \frac{\ln(\alpha_2)}{\alpha_3}; \\
\alpha_3^2 U_{i, \text{rp}1} U_{i, \text{rp}2} - U_{i, \text{rp}1} + \sum_{n=1}^{\infty} \frac{-1^{n+1} \cdot \alpha_2^n}{n^2 \cdot \exp(\alpha_3 U_{i, \text{rp}2} n)} - \\
\left\{ \begin{array}{l}
\frac{\ln^2(\alpha_2)}{2} + \sum_{n=1}^{\infty} \frac{\mathbb{1} \cdot \exp(\alpha_3 U_{i, \text{rp}1} n)}{n^2 \cdot \alpha_2^n}, U_{i, \text{rp}1} < \frac{\ln(\alpha_2)}{\alpha_3}; \\
\sum_{n=1}^{\infty} \frac{\mathbb{1}^{n+1} \cdot \alpha_2^n}{n^2 \cdot \exp(\alpha_3 U_{i, \text{rp}2} n)}, U_{i, \text{rp}1} > \frac{\ln(\alpha_2)}{\alpha_3};
\end{array} \right\}, U_{i, \text{rp}2} > \frac{\ln(\alpha_2)}{\alpha_3}; \\
0 \text{ в противном случае,}
\end{array} \right. \quad (4)$$

где $\beta = -\ln\{1 + \alpha_2 f_{o_{k,i}} \cdot \exp(-\alpha_3 \cdot U_{i, \text{rp}2})\} + \ln\{1 + \alpha_2 f_{o_{k,i}} \cdot \exp(-\alpha_3 \cdot U_{i, \text{rp}1})\}$, а $U_{i, \text{rp}1}$ и $U_{i, \text{rp}2}$ – границы участка x КМ $_i$. Наконец, для упрощения вычислений формулируется достаточное условие отсутствия маскировки: $K_x(k)$ не маскируется, если его уровень не меньше уровня КМ любого другого $K_x(i)$ на частоте f_k . В соответствии с этими результатами и в предположении независимости $K_x(k)$ получаем выражение для оценки искомой вероятности:

$$G_k = 1 - P\{U_k \geq A(f_k)\} \cdot \prod_{i=n_1}^{n_2} \left[1 - \sum_x g_{k,i}^{(x)} \right], \quad (5)$$

где $P\{U_k \geq A(f_k)\}$ – вероятность отсутствия маскировки $K_x(k)$ абсолютным порогом слышимости $A(f_k)$; n_1 и n_2 – соответственно, наименьший и наибольший номера $K_x(i)$, удовлетворяющие условию $U_{\text{км}}^{(3)}(f_{o_{k,i}}, D_i) > A(f_k)$; $g_{k,i}^{(x)}$ – условная вероятность маскировки $K_x(k)$ участком x КМ $_i$, определяемая в соответствии с известным выражением.

Помимо модуля, $K_x(k)$ характеризуются фазовым сдвигом. Ограничение разрядности модуля фазы φ_k ($k > 0$) на величину $\Delta\varphi_k$ приводит к изменению частоты f_k соответствующей компоненты \hat{s}_k нового спектра на величину $\Delta_{f,k}$. Причем эти изменения происходят на стыках спектральных выборок. Такой характер изменения можно интерпретировать как частотную модуляцию $K_x(k)$ с частотой модуляции $f_{\text{мод}} = 1/T_{\text{п}}$ и девиацией частоты $\Delta_{f,k}$. Это позволяет сопоставить относительное изменение фазы $\Delta\varphi_k / \varphi_k$ с индексом частотной модуляции $\Delta_{f,k} / f_{\text{мод}}$. Но тогда естественным критерием ограничения разрядности модуля фазы является минимально ощущаемая на слух девиация частоты δf . В результате анализа известных экспериментальных данных автором с высокой точностью (коэффициент детерминации $R^2 = 0,995$) получено соответствующее аналитическое выражение:

$$\delta f(f_k, f_{\text{мод}}) = K1(f_{\text{мод}}) \cdot 10^{-4} \cdot f_k^{e/2} + K2(f_{\text{мод}}) / \lg(f_k), \quad (6)$$

где коэффициенты $K1(f_{\text{мод}})$ и $K2(f_{\text{мод}})$ подбираются эмпирически.

Пусть теперь $|\varphi_k|$ некоторого немаскируемого $K_x(k)$ текущей спектральной выборки представлен $q_{\varphi,k}$ значащими двоичными разрядами. В результате округления $q2_{\varphi,k}$ младших разрядов обнуляются, вследствие чего достаточно передать только $q1_{\varphi,k} = q_{\varphi,k} - q2_{\varphi,k}$ старших разрядов. При этом по правилам округления $|\varphi_k|$ увеличится или уменьшится не более чем на величину $\Delta = \text{bin}(q_{\varphi,k} - q1_{\varphi,k}) / 2$, где

$\text{bin}(x) = 2^x$. Согласно принятой интерпретации округления отношение $\Delta/\text{bin}(q_{\varphi,k}) = \text{bin}(-q_{1,\varphi,k} - 1)$ определяет индекс частотной модуляции. Тогда искажения, вызванные округлением $|f_k|$, не будут ощущаться на слух, если это отношение не превышает δf . Отсюда

$$q_{1,\varphi,k} = \lceil \log_2[f_{\text{мод}}/\delta f(f_k, f_{\text{мод}})] - 1 \rceil,$$

где $\lceil x \rceil$ – наименьшее целое, не меньшее x .

Заметим, что округление $K_x(0)$ приводит к дополнительному фазовому сдвигу компонент нового спектра на стыках выборок на величину, по модулю не более $\Delta\varphi_0/2$. Однако если это не больше наименьшей допустимой степени округления модуля фазы, подобные искажения на слух ощущаться не будут. Следовательно,

$$q_{1_0} = \max_{k=r}^{N-1} q_{1,\varphi,k},$$

где r – номер первого

коэффициента рабочей полосы. Наконец, поскольку ДКП чувствительно к фазовым сдвигам, для всех $k > 0$ $q_{1_k} = \lceil \log_2(1/m_k) - 1 \rceil$, если $m_k < \delta f(f_k, f_{\text{мод}}) \cdot T_{\text{п}}$ и $q_{1_k} = q_{1,\varphi,k}$ в противном случае.

Метод оценки распределения ДСО для ЗС любого типа и при любых F_d и D в общем виде приведен на рис. 5. Как показывают расчеты, зависимость распределения ДСО от F_d и D слабая. Кроме того, в среднем q_{1_k} для ДКП по сравнению с ДПФ-Х практически в 2 раза меньше.

В третьей главе разрабатывается метод прореживания спектра и формулируются условия его максимальной эффективности.

Поскольку $2 \cdot \delta f$ является градацией раздражения и определяет ширину еще различимой на слух ступени частоты, $K_x(k)$ следует рассматривать как тональную компоненту ЗС, если $\delta f_k < \Delta f$ для ДПФ-Х и $\delta f_k < 2 \cdot \Delta f$ для ДКП, где Δf – разрешающая способность ДОП по частоте. В противном случае интерпретация δf как градации раздражения теряет смысл, что характерно для шума. На том же основании можно сформулировать условие разделения коэффициентов ДКП на два потока: $\delta f_k < 3 \cdot \Delta f$.

Как известно, для тональных компонент характерно дополнительное смещение $a_{m,i}$ на величину $\Delta_{\text{см}} = 10 \dots 12$ дБ. Кроме того, вследствие неравномерности округления фазы на последовательности выборок, в новом спектре на соответствующих частотах возможны биения. Но только при совокупности следующих обстоятельств: 1) $K_x(k)$ отображает тональную компоненту, 2) $k = 2m$ и/или $k = 3n$, 3) U_m и/или U_n превышают $U_{\text{кр}}$ (соответствует $f_{0\text{кр}}$ пересечения всех участков правой ветви КМ), 4) $K_x(m)$ и/или $K_x(n)$ не маскируются. При этом $K_x(k)$ маскируется, если $U_k < (U_{\text{км}} - \Delta_{\text{км}})$ дБ, где $\Delta_{\text{км}} = 8 \dots 10$ дБ.

На этих основаниях разрабатывается метод прореживания коэффициентов ДОП, обобщенная алгоритмическая схема которого представлена рис. 6, где

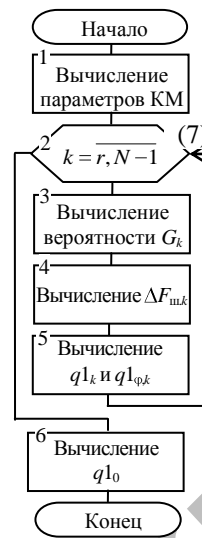


Рис. 5. Метод оценки q_{1_k}

ЛМК – локально маскирующий коэффициент очередной участок порога слышимости рассматриваются слектировки \hat{s}_k .

1. $k = 2m$. Так как событие отсутствия в случае биений, обусловленных \hat{s}_m отсутствием (событие \bar{B}_m) независимы и P_k отсутствия маскировки $K_x(k)$

$$P_k = P\{U_k \geq A(f_k)\} \cdot [P(OM_k) \cdot P(B_m) + P(OM_k) \cdot P(\bar{B}_m)] \quad (8)$$

где $P(OM_k) = \prod_{i=m}^{n_2} [1 - \sum_x g_{k,i}^{(x)}]$, но в первом

а во втором $U_m \leq U_{кр}$.

2. $k = 3n$. Здесь P_k определяется граница области возможно маскирования ограничена n .

3. $k = 2m$ и $k = 3n$. В этом случае

$$P_k = P\{U_k \geq A(f_k)\} \cdot [P(OM_k) \cdot P(B_m) + P(OM_k) \cdot P(\bar{B}_m) \cdot P(B_n) + P(OM_k) \cdot P(\bar{B}_m) \cdot P(\bar{B}_n)] \quad (9)$$

где $P(OM_k)$ первого слагаемого вычисляется аналогично первой ситуации, второго – второй ситуации, а третьего – при $(U_m, U_n) \leq U_{кр}$.

4. $k \neq 2m$ и $k \neq 3n$. Здесь $P_k = 1 - G_k$.

Относительно $P(B_j)$ достаточно заметить, что с вероятностью $1 - P(B_j)$ компонента \hat{s}_j не порождает биений, если она либо маскируется, либо по уровню не превышает $U_{кр}$. Отсюда сразу следует: $P(B_j) = P_k - u_{кр}/d_j$.

Теперь относительно числа немаскируемых коэффициентов $N_{нм}$ спектральной выборки установим пороговую величину $N_{п}$. Потребуем, чтобы средний интервал времени между событиями $N_{нм} \geq N_{п}$ был не меньше времени непрерывного звучания $T_{нз}$. Тогда $P\{N_{нм} < N_{п}\} > 1 - T_{п}/T_{нз}$. Усредняя эту вероятность по длине выборки, получаем условие оценки $N_{нм}$:

$$\sum_{l=0}^{N_{нм}-r} C_l^{N_{нм}-r} \cdot \bar{P}^l \cdot (1 - \bar{P})^{N_{нм}-r-l} > 1 - T_{п}/T_{нз}, \quad (10)$$

где C_x^y – число сочетаний из y по x , а $T_{нз} = 20$ мин для музыкальных сигналов (симфоническое произведение) и $T_{нз} = 9$ с для речевых (средний интервал между естественными паузами).

Для сравнения возможностей ДПФ-Х и ДКП в сжатии ЗС используется относительный коэффициент сжатия

$$\mu = \left[\bar{q}1_k^{(ДПФ-Х)} F_d + 3 \right] \cdot N_{нм}^{(ДПФ-Х)} F_d \left/ \left[\bar{q}1_k^{(ДКП)} F_d + 1 \right] \cdot N_{нм}^{(ДКП)} F_d \right., \quad (11)$$

где $\bar{q}1_k$ – усредненное на выборке распределение $q1_k$. Кроме того, учтены знаковые разряды, а для ДПФ-Х и минимально возможное количество дополнительных бит, необходимых для кодирования фазы. Как показывают расчеты $\mu > 1$ для любых F_d . Следовательно, в высококачественных системах сжатия цифровых аудиоданных целесообразным является ДКП.

Применительно к задаче оценке условий повышения эффективности сжатия

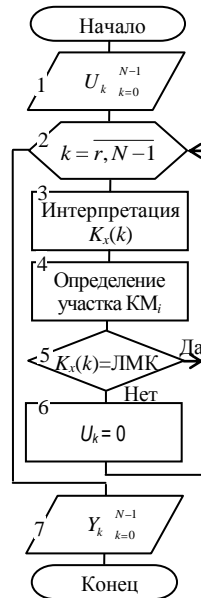


Рис. 6. Метод прореживания

коэффициент (определяет слышимости). С целью оценки его слышимости рассматриваются ситуации мас-

кировки (ОМ_k) \hat{s}_k (событие B_m), и при их несовместности, вероят-

$$P(OM_k) \cdot P(\bar{B}_m)],$$

слагаемом $U_m \in [U_{кр}, D_m]$,

аналогично (9), но нижерассматриваемых коэффициентов

ЗС с прореживанием спектра выражение (12) примет вид:

$$\mu = \overline{q1}_k D_{ст}, F_{д,ст} \cdot N_{нм} D_{ст}, F_{д,ст} / \overline{q1}_k D, F_{д} \cdot N_{нм} D, F_{д}, \quad (12)$$

где $D_{ст}$ и $F_{д,ст}$ – стандартные для данной системы D и $F_{д}$, соответственно.

Степень значимости D и $F_{д}$ схематично показана на рис. 7. При этом, как показывают расчеты, за счет изменения их значений можно повысить эффективность сжатия ЗС с прореживанием спектра в 1,5 раза, причем без ухудшения качества субъективного восприятия сжатого звука.

Наконец определяется

зона частот F_p , где про-

Суть соответствующего

последовательном

сторону верхних частот

$N^{(n)} / N^{(n-1)} > N_{нм}^{(n)} / N_{нм}^{(n-1)}$, где

ет номер итерации. Начиная с этого момента, скорость уменьшения области вышает скорость уменьшения $N_{нм}$. Следовательно, в области остались наиболее часто маскируемые коэффициенты. Результаты расчетов показывают, что реживание целесообразно в 90 % F_p .

В четвертой главе с помощью компьютерного моделирования и субъективно-статистического анализа проводится оценка степени достоверности полученного распределения $q1_k$ и предположения о биениях, а также проверка целесообразности применения прореживания спектра. Критерием достоверности является качество ЗС на выходе психоакустической модели, а критерием целесообразности – эффективность и качество ЗС на выходе алгоритма кодирования. В качестве тестовых фрагментов использовалась монофоническая дикторская речь и стереофоническая музыкальная фраза в полосе частот до 10 и 20 кГц, соответственно. Оба фрагмента записаны при $F_{д} = 44,1$ кГц с разрешением 16 бит/отсчет.

Приводится описание используемых в моделировании средств и условий проведения субъективной оценки качества сжатого звука. Во всех моделях спектр сигналов представлен коэффициентами ДКП.

На рис. 8 показаны теоретическое $q1_T(f)$ и модельное $q1_M(f)$ распределения $q1_k$ для музыкального тестового фрагмента. Суть методики моделирования соответствующих искажений заключается в уменьшении $q1_k$ в одной из ступеней $q1_M(f)$ и экспертной оценке качества звучания. Результаты экспериментов, ограниченные уменьшением на 1 бит только второй и только третьей ступени, приведены на рис. 9. Таким образом, зависимости $q1_T(f)$ и $q1_M(f)$ хорошо согласуются по форме, причем $q1_M(f) \leq q1_T(f)$. Тем самым подтверждается существование в ДКП фазовых искажений, а $q1_T(f)$ можно считать достаточной для любых типов и жанров ЗС.

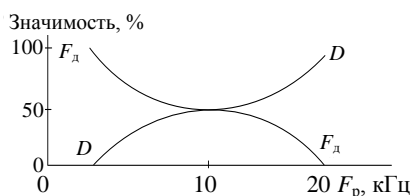


Рис. 7. Значение D и $F_{д}$ в повышении μ

область рабочего диапа-
реживание имеет смысл.
метода заключается в
нии исходной выборки в
до выполнения условия
верхний индекс

Для проверки возможности предлагается, изменяя $\Delta_{км}$, экспертную оценку качества делирования (рис. 10) погласиться с данной зой.

Для апробации реживания предлагает- модификация алго- ритма с поблочно пла- той (*Dolby AC-3*, *MPEG-2 AAC*). Коэф- фициенты спектраль- ной ки объединяются в бло- знаку одинакового $q1_k$. прореживание не имеет руются как обычно с мощностью порядка

ство нулевых старших разрядов для наибольшего по модулю коэффициента блока) и мантисс (ограниченное число старших разрядов коэффициентов блока). В остальных блоках только для немаскируемых коэффициентов помимо порядков и мантисс кодируются расстояния между ними. При этом теоретическая скорость цифрового потока ($V_{цип}$) для музыкального и речевого тестового фрагмента составляет 107,6 и 101 кбит/с, соответственно.

В результате моделирования получена $V_{цип}$ для музыкального фрагмента 105,5 кбит/с при среднем балле качества -0,19, а для речевого – 99,5 кбит/с при среднем балле качества -0,23. Таким образом, при достаточно высоком качестве субъективного восприятия звука за счет прореживания $V_{цип}$ относительно наиболее перспективного на сегодняшний день алгоритма кодирования MPEG-2 AAC (128 кбит/с при дополнительном статистическом уплотнении и среднем балле качества -0,17) снизилась в 1,2 раза для музыкального фрагмента и в 1,3 для речевого фрагмента.

Хорошее согласование теоретических и экспериментальных результатов позволяет сделать прогноз эффективности применения прореживания при нестандартных F_d и D . Как показывают расчеты, в этих условиях можно обеспечить $V_{цип}$ не более 96,7 кбит/с ($\mu = 1,3$) для музыкальных сигналов, 75,7 кбит/с ($\mu = 1,7$) для широкополосных речевых сигналов и 31,5 кбит/с ($\mu = 2$) для телефонии. И это без дополнительного статистического уплотнения и при высоком качестве субъективного восприятия звука.

В заключении приведены основные результаты работы:

1. На основе психофизиологических свойств слуха сформулированы требования, позволяющие обоснованно установить ДОП, перспективные для систем высококачественного сжатия цифровой аудиоинформации.
2. Разработан метод оценки распределения ДСО коэффициентов гармонических ДОП в пределах спектральной выборки, позволяющий рационально рас-

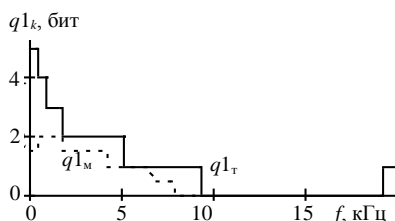


Рис. 8. Распределение $q1_k$



Рис. 9. Экспертиза распределения $q1_k$



Рис. 10. Экспертиза возможных биений

ности биений
всякий раз проводить
ва звука. Результаты мо-
звоняют со-
те-

про-
ся

вающей запя-

фи-
выбор-

ки по при-
Блоки, где
смысла, коди-

по-
че-

пределить биты при кодировании сжатого спектра с учетом достаточно высокого качества субъективного восприятия сжатого звука.

3. Получены аналитические выражения для оценки вероятности маскировки коэффициентов гармонических ДОП при различной их интерпретации и использовании всех ветвей кривой маскировки.

4. Разработан алгоритм прореживания спектра, позволяющий за счет изменения D и F_d дополнительно повысить эффективность компрессии цифровых аудиоданных в 1,5 раза при достаточно высоком качестве субъективного восприятия сжатого звука.

5. Разработан метод оценки эффективной для прореживания области рабочего диапазона частот, расширяющий перспективу развития новых более эффективных методов кодирования сжатых цифровых аудиоданных.

6. Разработаны методики экспериментальной обработки цифровой аудиоинформации с прореживанием спектра средствами компьютерного моделирования, позволяющие оценить эффективность новых алгоритмов сжатия ЗС.

7. Результаты проведенных исследований могут быть полезны при разработке новых рациональных с точки зрения реализационной сложности и высокоэффективных кодеков для высококачественных систем цифровой обработки и передачи аудиоданных. Применение их в лабораторных условиях относительно наилучшего алгоритма кодирования MPEG-2 AAC практически при тех же качественных показателях, стандартных параметрах цифровых аудиоданных позволило снизить скорость цифрового потока в 1,2 раза.

Приложение содержит протокол экспертной комиссии субъективных оценок качества сжатых ЗС и акты использования результатов диссертации.

Публикации по теме диссертации

1. Стефанов, М. А. Эффективные дискретные ортогональные преобразования в компрессии цифровых аудиоданных / М. А. Стефанов. // «Инфокоммуникационные технологии». – 2009. – т. 7. – № 4. – С. 17-22.

2. Акчурин, Э.А. Энергетический параметр высококачественной эффективной компрессии цифровых аудиоданных / Э.А. Акчурин, А.М. Стефанов, М. А. Стефанов. // «Инфокоммуникационные технологии». – 2009. – т. 7. – № 2. – С. 82-87.

3. Стефанова, И.А. Оптимизация временного параметра эффективности сжатия звуковых сигналов / И.А. Стефанова, М.А. Стефанов. // 6 Междунар. науч.-техн. конф. «Проблемы техники и технологий телекоммуникаций»: сб. докладов. – Уфа, 2005. – С. 38-40.

4. Стефанов, М.А. Оценка допустимой степени округления фазы спектральных компонент звуковых сигналов / М.А. Стефанов. // VII Междунар. науч.-техн. конф. «Проблемы техники и технологий телекоммуникаций»: материалы конференции. – Самара, 2006. – С. 171-173.

5. Стефанов, М.А. Дискретные ортогональные преобразования как инструмент сжатия звуковых сигналов / М.А. Стефанов, Э.А. Акчурин. / XIV Российская науч. конф.: тез. докладов. – Самара, 2007. – С. 147-148.

6. Стефанов, М.А. Особенности округления спектральных компонент звуковых сигналов / М.А. Стефанов. // XV Российская науч. конф.: тез. докладов. –

Самара, 2008. – С. 185.

7. Стефанов, М.А. Оценка ширины полосы концентрации энергии коэффициентами ДОП / М.А. Стефанов. // IX Междунар. науч.-техн. конф. «Проблемы техники и технологий телекоммуникаций»: материалы конференции. – Казань, 2008. – С. 140-141.

8. Стефанов, М.А. Реализация округления коэффициентов дискретных ортогональных преобразований / М.А. Стефанов, А.М. Стефанов, И.А. Стефанова. // XV Российская науч. конф.: тез. докладов. – Самара, 2008. – С. 188.

9. Стефанов, М.А. К выбору способа нормировки коэффициентов ДОП при цифровой обработке звукового сигнала / М.А. Стефанов. // XVI Российская науч. конф.: тез. докладов. – Самара, 2009. – ч.2 – С. 12.

10. Стефанов, М.А. Особенности цифровой обработки звуковых сигналов в системе MATLAB / М.А. Стефанов. // XVI Российская науч. конф.: тез. докладов. – Самара, 2009. – ч.2 – С. 13.

11. Стефанов, М.А. Оценка вероятности маскировки спектральных компонент звуковых сигналов / М.А. Стефанов. // Электроника и информационные технологии. – 2009 Специальный выпуск (V Международная конференция "Методы и средства управления технологическими процессами" 19-21 ноября 2009. г. Саранск) – 2009. – <http://fetmag.mrsu.ru/2009-2/pdf/Stefanov.pdf>. – 0420900067/0015.