

# MPEG КАК ПРЕДВЕСТНИК ЭРЫ ИНТЕРАКТИВНОГО ТЕЛЕВИДЕНИЯ

М. Мейтин

**Магия движущейся картинки с незапамятных времен очаровывает людей. В эпоху мультимедиа появилась возможность оцифровки видеоинформации и ее обработки в реальном времени. Однако решение подобных задач требует огромных ресурсов. Например, одна минута цифрового видеосигнала с разрешением, сопоставимым с VHS, и цветопередачей true color (миллионы цветов), займет (288 x 358) пикселей x 24 бита x 25 кадров/с x 60 с = 442 Мбайт. Таким образом, на носителях, применяемых в современных ПК, т.е. на CD-ROM или жестком диске, сохранить полноценное по времени видео не удастся. Единственно возможное решение проблемы – использование алгоритмов сжатия информации.**

MPEG (Moving Picture Expert Group) – экспертная группа ISO, действующая в направлении разработки стандартов кодирования и сжатия видео- и аудиоданных. Официальное название группы – ISO/IEC JTC1 SC29 WG11. Аббревиатурой MPEG часто обозначают и стандарты, разработанные этой группой. На сегодняшний день из них известны следующие.

**MPEG-1** – разработан для записи синхронизированных видеоизображений (обычно в формате SIF, 288 x 358) и звукового сопровождения на CD-ROM с учетом максимальной скорости считывания около 1,5 Мбит/с. Качественные параметры видеоданных, обработанных MPEG-1, во многом аналогичны обычному VHS-видео. Поэтому формат применяется в первую очередь там, где неудобно или непрактично использовать стандартные аналоговые видеоносители.

**MPEG-2** – для обработки видеоизображения, соизмеримого по качеству с телевизионным при пропускной способности системы передачи данных в пределах от 3 до 15 Мбит/с (в профессиональной аппаратуре используют потоки скоростью до 50 Мбит/с). На технологии, основанные на MPEG-2, переходят многие телеканалы. Сигнал, сжатый в соответствии с этим стандартом, транслируется через телевизионные спутники, используется для архивации больших объемов видеоматериала.

**MPEG-3** – предназначался для использования в системах телевидения высокой четкости (high-definition television, HDTV) со скоростью потока данных 20-40 Мбит/с, но позже стал частью стандарта MPEG-2 и отдельно теперь не упоминается. Кстати, формат MP3, который иногда путают с MPEG-3, предназначен только для сжатия аудиоинформации, и полное название MP3 звучит как MPEG-Audio Layer-3.

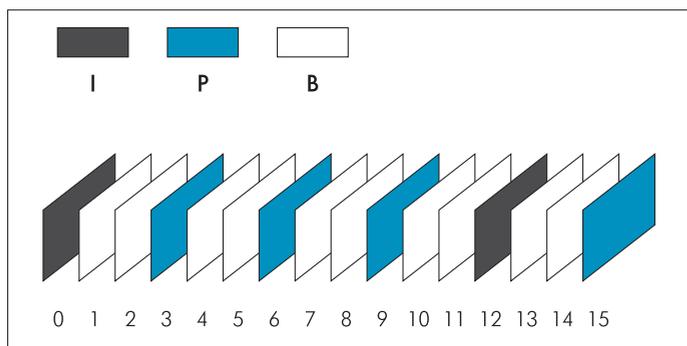
**MPEG-4** – задает принципы работы с изображением и звуком для трех областей: интерактивного мультимедиа (включая продукты, распространяемые на оптических дисках и через сеть), графических приложений (синтетического контента) и цифрового телевидения.

Базовым объектом кодирования в стандарте MPEG [1] является кадр телевизионного изображения. Стандарт предусматривает четыре разных типа кадров:

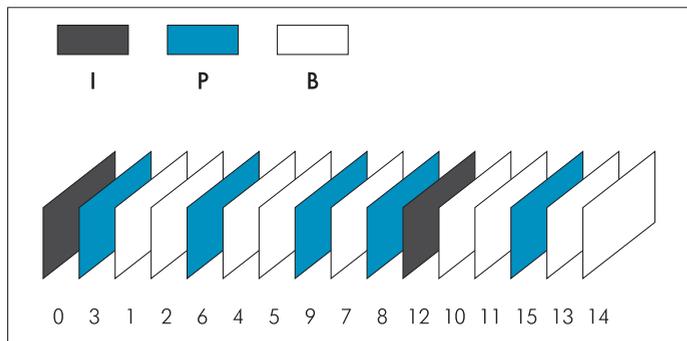
- I-кадры (Intra) – исходные кадры, которые кодируются без учета взаимосвязи с другими кадрами;
- P-кадры (Predicted) – предсказываемые кадры, использующие информацию предшествующих кадров;
- B-кадры (Bidirectionally predicted) – кадры двунаправленного предсказания, кодируемые с использованием данных из предыдущих и следующих кадров;
- D-кадры, которые служат исключительно для средств быстрого поиска.

Поскольку в большинстве фрагментов фон изображения остается достаточно стабильным, а действие происходит только на переднем плане, сжатие начинается с создания исходных кадров. Эти кадры кодируются с применением внутрикадрового сжатия по алгоритмам, аналогичным используемым в JPEG. Кадр разбивается на блоки 8x8 пикселей. Над каждым блоком производится дискретно-косинусное преобразование (ДКП) с последующим квантованием полученных коэффициентов. Вследствие высокой пространственной корреляции яркости между соседними пикселями изображения ДКП приводит к концентрации сигнала в низкочастотной части спектра, который после квантования эффективно сжимается с использованием кодирования кодами переменной длины. Обработка предсказываемых (Predicted) кадров производится с использованием предсказания вперед по предшествующим исходным или предсказываемым кадрам. Кадр разбивается на макроблоки 16x16 пикселей. Каждому макроблоку ставится в соответствие наиболее похожий участок изображения из опорного кадра, сдвинутый на радиус-вектор перемещения (Motion Vector). Эта процедура называется анализом и компенсацией движения. При этом степень сжатия для предсказываемых кадров в три раза больше, чем для ис-

*Из всех искусств для нас  
важнейшим является кино.  
В.И. Ленин*



**Рис.1. Последовательность кадров при их показе**



**Рис.2. Последовательность кадров при их передаче**

ходных. В зависимости от характера видеоизображения кадры двунаправленной интерполяции (Bi-directional Interpolated) кодируются следующими способами: предсказание вперед; обратное предсказание с компенсацией движения – используется, когда в кодируемом кадре появляются новые объекты изображения; двунаправленное предсказание с компенсацией движения – при резкой смене сюжета или при высокой скорости перемещения элементов изображения. С двунаправленными кадрами связано наиболее глубокое сжатие видеоданных, но, поскольку высокая степень сжатия снижает точность восстановления исходного изображения, двунаправленные кадры не используются в качестве опорных. Если бы коэффициенты ДКП передавались точно, восстановленное изображение

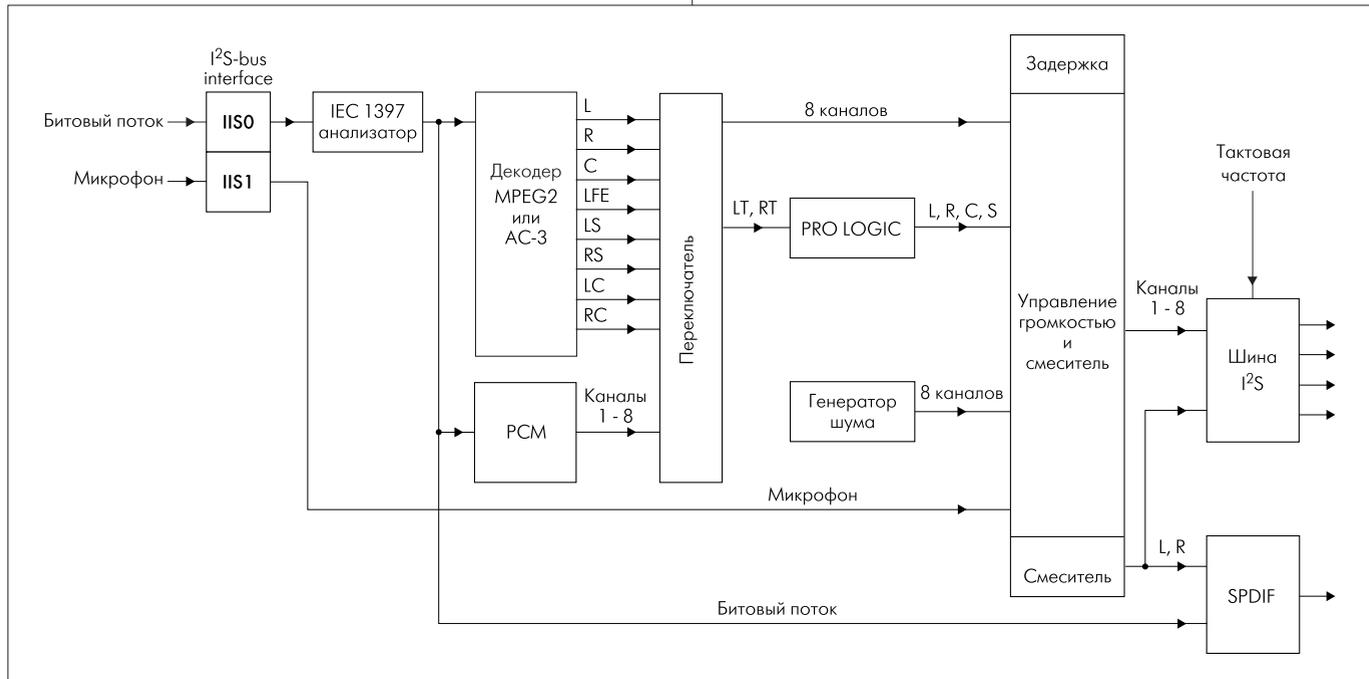
полностью совпадало бы с исходным. Однако ошибки восстановления коэффициентов ДКП, связанные с квантованием, приводят к искажениям изображения. Чем грубее производится квантование, тем меньший объем занимают коэффициенты и тем сильнее сжатие сигнала, но и тем больше визуальных искажений.

Типичная последовательность кадров (рис.1) содержит один I-кадр на каждые 10-15 кадров и два B-кадра между I- и P-кадрами. Однако передаются кадры в другой последовательности (рис.2), поскольку для успешного декодирования необходимо сначала получить информацию об I- и P-кадрах.

Стандартом MPEG предусматривается три формата кодирования звука: Layer I, Layer II и Layer III [2]. Последний достигает наибольшего сжатия, однако требует больше ресурсов на кодирование. Принципы кодирования основаны на том, что в звуке передается много избыточной для человеческого уха информации. Например, сильный звук с частотой 1000 Гц заглушает слабый звук с частотой 1100 Гц. Кроме того, чувствительность уха будет ослаблена за 5 мс до возникновения сильного звука и на 100 мс после него. Таким образом, психоакустическая модель, используемая в MPEG, разбивает весь частотный спектр на части, в которых уровень звука считается одинаковым, а затем удаляет звуки, не воспринимаемые человеком. В Layer III части разбитого спектра самые маленькие, что обеспечивает самое лучшее сжатие. Существует множество чипсетов, предназначенных для кодирования и декодирования звука, в частности декодер SAA250H от Philips Semiconductors (рис.3) и STA015 корпорации ST Microelectronics (рис.4) [3,4].

Синхронизация и объединение звука и видео осуществляются с помощью системного потока (System Stream), как показано на рис.5. Системный поток состоит из двух слоев: системного, содержащего временную и другую информацию для разделения и синхронизации видео- и аудиопотоков, и компрессионного, содержащего видео- и аудиопотоки. Видеопоток содержит заголовок и несколько групп картинок. Аудиопоток состоит из заголовка и нескольких звуковых кадров. Для синхронизации видео- и аудиопотоков в системном потоке используется таймер с частотой 90 кГц.

По мнению многих специалистов, в ближайшее время произойдет рост рынка аппаратных декодеров MPEG-2 [5]. Они выпускают-



**Рис.3. Блок-схема аудиодекодера SAA250H**

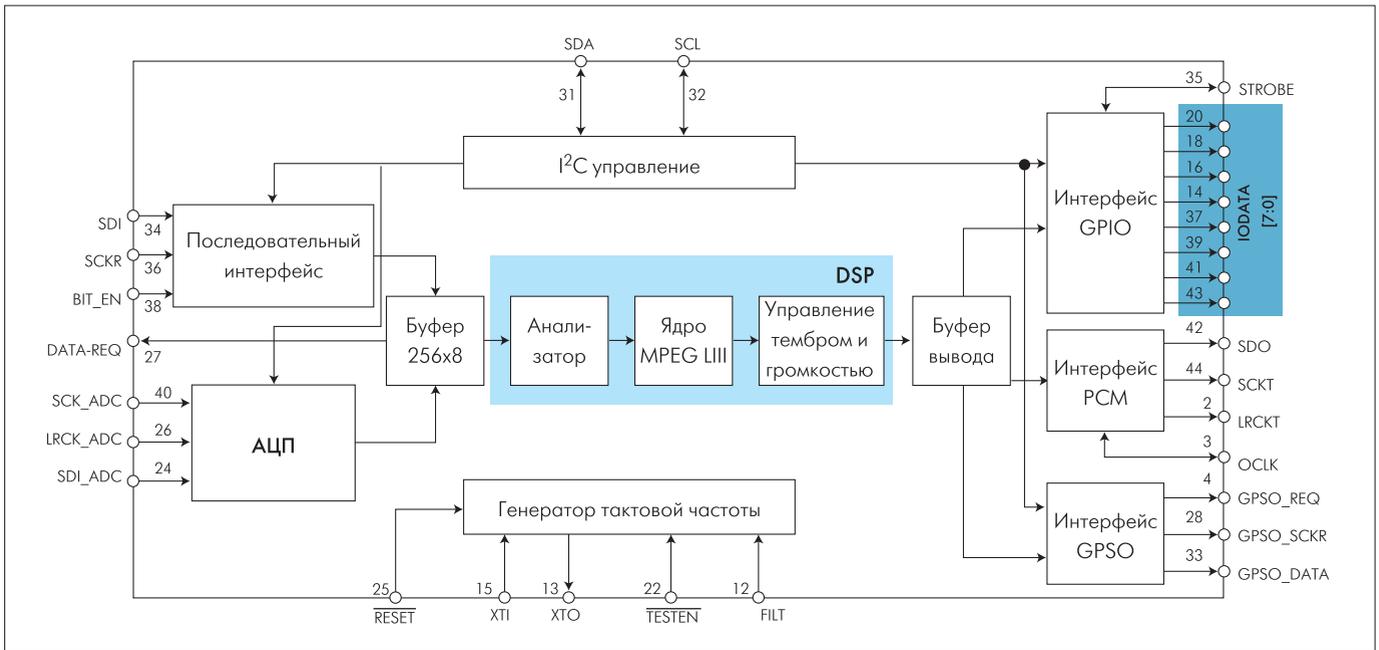


Рис.4. Блок-схема аудиодекодера STA015

ся уже несколько лет и находят применение во многих устройствах записи и воспроизведения видео- и звуковых CD, плейерах DVD, не говоря уже о практически всех современных графических акселераторах. Среди фирм-производителей декодеров MPEG-2 – компании Stream Machine и iCompression. Главный продукт Stream Machine – микросхема SM2210, реализующая полный набор функций кодирования/декодирования MPEG-2. Она совместима со стандартами PAL и NTSC и ориентирована на использование как в потребительских устройствах (видеомагнитофоны, цифровые камеры, DVD-плейеры), так и в дополнительных платах для ПК. Компания iCompression, в свою очередь, предлагает чипсет из двух микросхем – ITVC10 и ITVC12. Компания STMicroelectronics также предлагает аппаратные декодеры MPEG-2 – Omega, ориентированные на рынок устройств DVD и цифровых телеприставок. Последний декодер компании – Sti5512 – был выпущен в ноябре прошлого года. Известный поставщик средств мультимедиа – ESS Technology – также не остался в стороне, выпустив декодер ES3308, который можно применять в устройствах на базе DVD и в телевизионных приставках.

Однако после появления стандарта MPEG-4 его предшественники превратились лишь в частные случаи кодирования одного из многочисленных типов данных [6]. Важность стандарта MPEG-4 невозможно переоценить – по существу он представляет собой единый концептуальный способ описания, представления и обработки мультимедийных данных. Фактически он задает правила организации объектно-ориентированной среды и имеет дело не просто с потоками и массивами данных, а с медиаобъектами. Объекты могут быть аудио-, видео-, аудиовизуальными, графическими (плоскими

и трехмерными), текстовыми. Они могут быть как естественными (записанными, отснятыми, отсканированными), так и синтетическими (искусственно сгенерированными). Примерами объектов могут служить неподвижный фон, видеоперсонажи отдельно от фона, синтезированная на основе текста речь, музыкальные фрагменты, трехмерная модель, которую можно двигать и вращать в кадре, и многое другое. Медиаобъекты могут быть потоковыми. Каждый из них имеет свой набор свойств, определяющих его положение на сцене, и операций, необходимых для декодирования потоковых данных. Сцена имеет свою систему координат, в соответствии с которой размещаются объекты. Звуковые объекты также имеют свои координаты в пространстве сцены, которые могут меняться для создания стерео- и surround-эффектов. Сцена описывается с помощью иерархической структуры, узлами которой являются объекты. Она динамически перестраивается по мере того, как узлы-объекты добавляются, удаляются или заменяются. Для описания объектов, классов объектов и сцен в MPEG-4 определен язык BIFS, являющийся расширением C++. С помощью BIFS можно анимировать объекты, менять их координаты, размеры, свойства, задавать поведение, реакции на воздействие пользователя, выполнять двухмерные построения и т.д.

Важнейшая особенность MPEG-4 – в том, что окончательная сборка сцены (с возможностью добавления разного рода геометрических преобразований, аудио- и видеоэффектов реального времени) происходит на приемном конце – в компьютере, телевизоре или телевизионной приставке пользователя. Таким образом, пользователь сам может формировать получаемое изображение, играя роль телережиссера. Более того, среди допустимых пользовательских команд – изменение точки наблюдения, удаление, добавление и перемещение объектов внутри сцены и многое другое. Конечно, такое воздействие должно быть разрешено создателем аудиовизуального потока информации. Команды пользователя могут быть обработаны в декодере или пересланы на передающую сторону.

Для кодирования видеoinформации стандарт предусматривает несколько отдельных блоков. MPEG-4 разрабатывался как способ передачи медиаданных по каналам с низкой пропускной способностью (4,8 – 64 Кбит/с). Согласно стандарту, эту задачу выполняет блок VLBV Core (Very Low Bit-rate Video), имеющий ухудшенные характери-

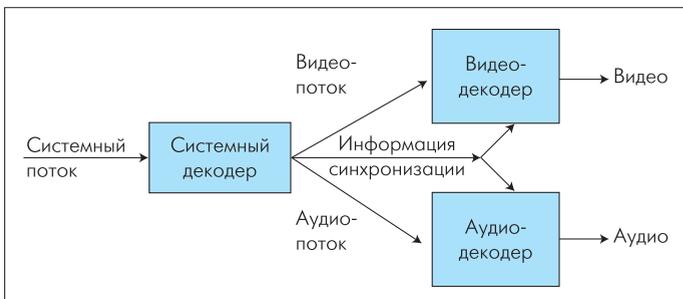


Рис.5. Схема синхронизации звука и видео



стики по разрешению (320x240 точек) и частоте кадров (15 кадр/с). Помимо эффективных и помехоустойчивых методов кодирования последовательности кадров VLBV реализует операции произвольного доступа к кадрам видеопоследовательности, а также быстрой перемотки видеоряда вперед и назад, что требуется для работы с базами видеоданных, хранящих видеоматериалы в низком разрешении (для целей быстрого поиска и оценки) и ссылки на соответствующие области хранения материалов вещательного качества.

Второй блок отвечает за работу с видео с большой скоростью потока, вплоть до вещательного качества по стандарту ITU-R 601, и обеспечивает функции, сходные с блоком VLBV.

Названные блоки обрабатывают обычные потоки видео и фактически включают в себя функциональность MPEG-1 и MPEG-2, а также кодирование текстов.

Блок, выполняющий функции, зависящие от контента, предназначен для обработки видео с произвольным силуэтом, например "вырезанного" силуэта диктора, и интерактивных манипуляций с ним. Помимо обычных методов кодирования – предсказания и компенсации движения – предусмотрены механизмы работы со "спрайтами" – неподвижными изображениями, которые передаются в декодер лишь однажды и всякий раз подставляются в нужное место кадра из специального спрайтового буфера. Механизм спрайтов позволяет значительно снизить объем передаваемых данных и обеспечивает большую гибкость при построении сцен. Например, поверх "живого" видеofона можно запустить различные объекты-спрайты (самолеты, автомобили) или наоборот, создать в качестве фона спрайт-панораму с "живыми" объектами поверх нее.

Этот же блок отвечает за масштабируемость видеообъектов – объекты кодируются таким образом, чтобы декодер имел возмож-

ность в случае ограничений пропускной способности сети или параметров самого декодера сохранять адекватность передачи контента за счет огрубления изображения (уменьшая частоту кадров, увеличивая зернистость). Вообще, для потоков предусмотрено до трех уровней зернистости. При кодировании неподвижных изображений и текстур в MPEG-4 применяется очень эффективный wavelet-алгоритм, обеспечивающий кодирование объектов произвольной формы, 11 уровней масштабируемости по разрешению, плавную масштабируемость по качеству картинки. Результирующий поток представляет собой "пирамиду" различных разрешений, и в приемнике картинка со временем "проявляется", улучшаясь настолько, насколько позволяет передающая среда.

В MPEG-4 предусмотрены инструменты и алгоритмы для работы не только с "живыми" видеообъектами, но и с объектами синтетическими, полученными с помощью средств компьютерной графики и анимации: двух- и трехмерными каркасными моделями и текстурами, которые накладываются на эти модели. Подобные объекты позволяют значительно сократить объем передаваемых данных, так как для их анимации достаточно передавать лишь необходимые параметры, а все остальное сделает декодер.

Особое место среди синтетических объектов стандарт уделяет анимации человеческого лица. В MPEG-4 установлены наборы параметров для задания особенностей лица, для его анимации, контрольные точки в полигональной сетке, ответственные за выражение эмоций. Кроме того, на построенную модель лица можно натянуть фото- или видеотекстуру лица конкретного человека.

Звуковая часть MPEG-4 также объектно-ориентирована. Аудиообъекты входят в структуру сцены и описываются на языке BIFS, что позволяет располагать источники звука в трехмерном пространстве

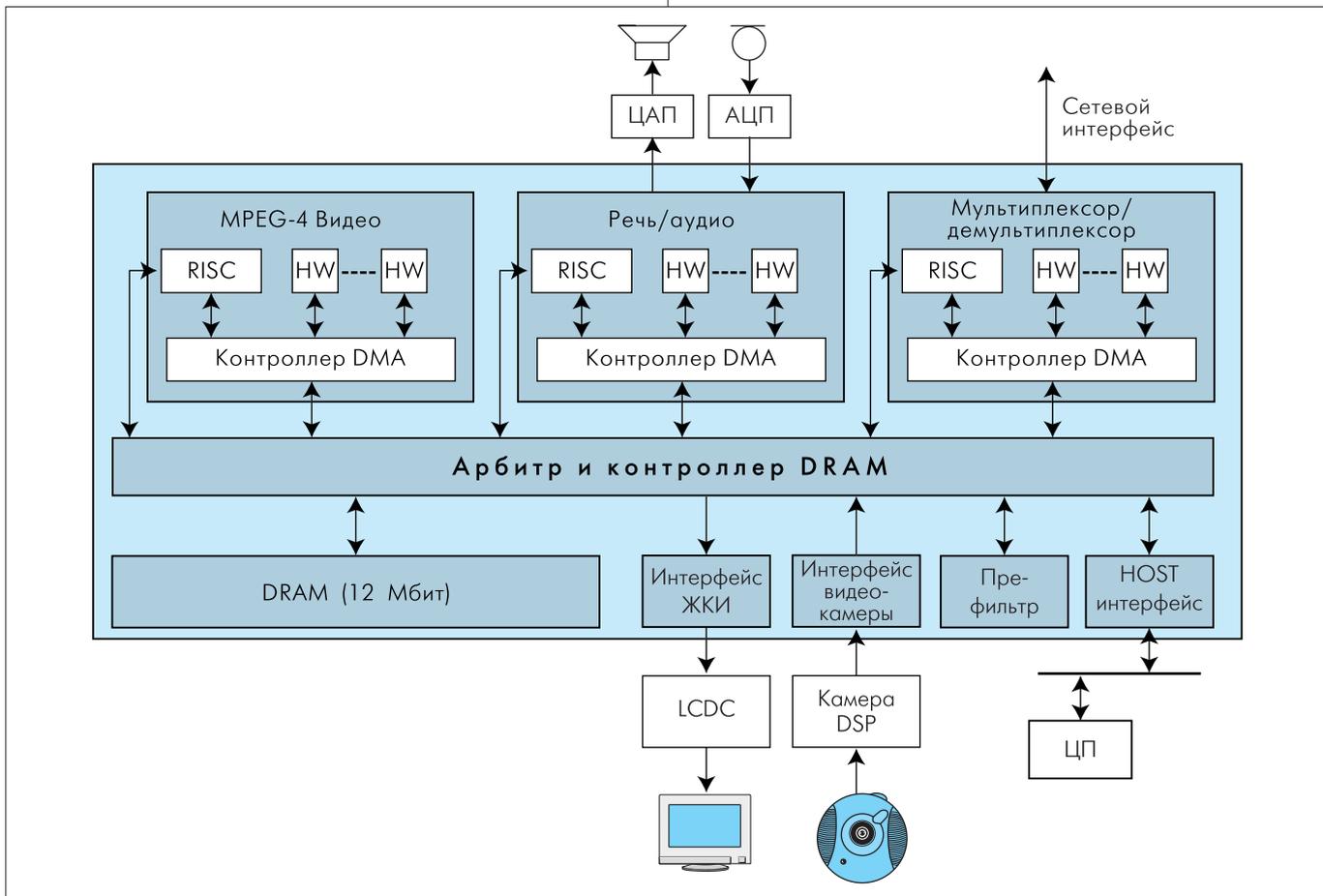


Рис.6. Блок-схема MPEG-4 декодера TC35273

сцены, управлять их характеристиками и применять к ним различные эффекты независимо друг от друга, перемещать источник звука при перемещении связанного с ним визуального объекта и т.п.

Для кодирования аудиообъектов MPEG-4 предлагает наборы музыкальных инструментов – как для живых звуков, так и для синтезированных. MPEG-4 устанавливает синтаксис двоичных потоков и процесс декодирования в терминах наборов инструментов. Это позволяет применять различные алгоритмы сжатия. Диапазон предлагаемых стандартом скоростей потока для кодирования живых звуков – от 2 до 128 Кбит/с и выше. Для кодирования речи предлагаются алгоритмы: HVXC (Harmonic Vector eXcitation Coding) – для скоростей потока 2–4 Кбит/с и CELP (Code Excited Linear Predictive) – для скоростей 4–24 Кбит/с.

MPEG-4 предлагает в качестве стандарта средства синтеза произвольных звуков и музыки. Этот подход, разработанный в MIT Media Lab и названный Structured Audio (SA) – “Структурированный звук”, определяет формат описания методов синтеза. Для этого введено два языка: SAOL (Structured Audio Orchestra Language) и SASL (Structured Audio Score Language). Как следует из названия, первый задает оркестр, а второй – то, что этот оркестр должен играть. Каждый инструмент оркестра представлен набором элементов цифровой обработки сигналов – синтезаторов и цифровых фильтров, которые все вместе и синтезируют нужный звук. С помощью SAOL можно запрограммировать практически любой нужный инструмент, природный или искусственный звук. Сначала в декодер загружается набор инструментов, а затем поток данных SASL заставляет этот оркестр играть, управляя процессом синтеза.

Корпорация Toshiba первой представила свой кодек MPEG-4 – однокристалльную микросхему TC35273, созданную по технологии

0,18 мкм [7]. В ее состав входят три устройства обработки сигналов: видеокодек, аудиокодек и мультиплексор/демультиплексор (рис.6). В каждом из них – 16-битный RISC-процессор и специализированный аппаратный акселератор для обеспечения высокой производительности и низкого энергопотребления.

Итак, перспективы применения MPEG-4 оказываются значительно более широкими, чем просто сжатие изображения и звука. Реально MPEG предложила и продолжает развивать революционный подход к передаче аудиовизуальной информации, безусловно оказывающий влияние на развитие информационных технологий и телекоммуникаций будущего. Благодаря MPEG человечество оказалось на пороге эры интерактивного телевидения. Сейчас MPEG разрабатывает новый стандарт – MPEG-7. Он будет обеспечивать описание различных типов мультимедийной информации для эффективного и быстрого ее поиска. Таким образом, можно не сомневаться, что объектно-ориентированному подходу обеспечено большое будущее.

#### ЛИТЕРАТУРА

1. An overview of MPEG compression algorithm. – Technical Note: SGS-Thomson Microelectronics.
2. Шурыгина В. Аудиоинтернет. Революция, или, скорее, эволюция? – ЭЛЕКТРОНИКА: НТБ, 2000, №4.
3. Philips Semiconductors SAA2505H Datasheet.
4. ST Microelectronics STA015 Datasheet.
5. Смалый А. Рынок декодеров MPEG-2: на пороге рассвета? ITC Online – <http://itc-ua.com>
6. <http://www.cselt.it/mpeg/standards/mpeg-4/mpeg-4.htm>
7. <http://www.toshiba.com>