DAT330 – Principles of Digital Audio
Cogswell Polytechnical College
Spring 2009

Week 10 – Class Notes

Desktop Audio

PC Buses and Interfaces

Internal and external peripherals can be interfaced to a host computer with any of several interconnections:

ISA (Industry Standard Association) local bus: found in IBM-PC computers. Has a data rate of 11 Mbyte/sec, which is limited for audio applications. For example, CD quality audio data stream, at 1.4 Mbps, along with sound effects and multiple voices would consume 40% of the bus.

EISA (Extended ISA): offers a data transfer rate of 32 Mbyte/sec, but is is also limited for audio applications.

PCI (Peripheral Component Interconnect): high performance local interconnection bus providing a maximum transfer rate of 132 Mbyte/sec, 32-bit pathways, good noise immunity, and an efficient integration of the computer's processor and memory with peripheral devices. It is lightly taxes with even multiple audio streams. Has lower latency than ISA and EISA. The PCI bus has replaced the ISA bus, and it is found on both IBM-PC and Macintosh computers.

ATA (Advanced Technology Attachment): also known as IDE (Integrated Drive Electronics). Designed as an expansion slot format for the PC providing 16-bit path width and speeds of 8.3 Mbps. Variations of the format, such as EIDE, ATA-2, or Ultra-ATA, can accommodate speeds up to 33 Mbps.

PCMCIA (Personal Computer Memory Card International Association): used on notebook computer expansion ports.

SCSI (Small Computer System Interface): high-speed data transfer protocol that allows multiple devices to access information over a common parallel bus (daisy-chaining).Devices are given a unique address or ID number. SCSI defines a number of variants, differing principally in data-path width and speed, as well as physical connectors. For example, the basic SCIS-1 width (number of bits in the parallel data path) is 8-bit (narrow). But 16- (wide) and 32-bit (very wide) versions can be used. SCSI-1 has a data transmission rate of 12.8 Mbps. Other speeds include Fast, Ultra, Ultra-2, and Ultra-3.

Because there are many interconnection problems found in PC's, the Plug and Play standard was devised. The operating system and system BIOS automatically configure jumper, IRQ, DMA address, SCSI ID's, and other parameters for the plug-in device.

IEEE 1394 “FireWire”

The 1394-1995 High Performance Serial Bus standard, based on the Apple Computer “FireWire” protocol, specifies a physical layer and controller for both a backplane bus and a serial data bus that allows inexpensive, general purpose, high-speed data transfer. This protocol is is a universal, platform-independent digital interface that can connect digital devices such as personal computers, audio/video devices, digital cable boxes and HDTV tuners, printers and scanners, digital video cameras and displays.

The IEEE 1394 cable standard defines three data rates: 98.304, 196.608, and 393.216; these rates are rounded to 100, 200, and 400 Mbps, and are referred to as S100, S200, and S400. The latter is also known as FireWire 400. Variations of the protocol allow for greater data rates, such as 800 Mbps (FireWire 800).

Devices can be connected point-to-point, or can be daisy-chained along lengths of cable. Connecting cables are usually made out of copper, however Cat5 and optical fiber cables can be used to extend the cable runs.

Up to 63 devices can be connected to a single local cluster before a bus bridge is needed, and up to 1023 clusters can be connected via data bridges. IEEE 1394 is “hot pluggable” meaning that the devices can be removed from a powered bus. And it is also a scalable architecture, so that several devices with multiple speeds can be connected on one bus. Although ultimately, the data transfer rate is determined by the slowest active node.

IEEE 1394 defines three layers: physical, link, and transaction. The physical layer defines the signals required by the IEEE 1394 bus. The link layer formats raw data (from the physical layer) into recognizable IEEE 1394 packets. The transaction layer presents packets (from the link layer) to the application.

When a connection is made, the bus automatically reinitializes the entire bus, recognizing the new device and integrating it with other networked devices. Similarly, upon disconnection the bus reconfigures itself.

Asynchronous transmission performs simple data transfer. The process is as follows: a chunk of data is transferred after acknowledgement that a previously transmitted chunk of data has been received. Timing cannot be predicted because of network demands, timing delivery is random. This is problematic for real-time audio/video data.

Isochronous transmission allows for high-bandwidth and low latency for data-intensive applications, such that synchronized data (audio-video) will be conveyed with small discrepancies so that they can be synchronized at the output.

The IEEE 1394 specification includes an “Audio and Music Data Transmission Protocol” (A/M protocol) that defines how real-time audio can be conveyed over IEEE 1394 using isochronous packets. Data types such as raw audio samples (up to 24-bit) and MIDI are supported. This protocol and multichannel versions provide sufficient bandwidth to convey DVD-Audio data streams of 9.6 Mbps. Copy protection methods are used to guard against the piracy of DVD-Audio signals.

The mLAN specification (developed by Yamaha) can be used to send multiple sample-accurate AES3 signals, raw audio, MIDI, and other control information over an IEEE 1394 bus. This protocol uses an isochronous transfer mode that ensures on-time delivery and also prevents collisions, and reduces latency and jitter. The device interconnection topology is similar to the IEEE 1394 specification and it is “hot pluggable”. Portion of mLAN was adopted by the IEEE 1394 specification.

The IEEE 1394b standard has throughputs of 800 and 1600 Mbps, and is also known as S800 and S1600. The first being also known as FireWire 800. It allows for daisy-chaining of different peripherals and cable lengths up to 800 m can be used for networks by using CAT5 or optical fiber cables. The IEEE 1394b ports have a different physical configuration from that seen in 1394.

Universal Serial Bus (USB)

Designed to replace the older computer serial and parallel I/O buses, to provide a faster, more user-friendly interconnection method, and to overcome the limitation of too-dew interrupts available for peripherals. All sorts of peripheral devices are candidates for USB. The original USB specification (USB 1.1) provides low-speed interconnection (12 Mbps), while the newer USB 2.0 specification provides data transfers up to 40-times faster than USB 1.1 (480 Mbps). USB 2.0 is fully compatible with USB 1.1 devices and uses the same connector, but USB 1.1 devices cannot operate at faster rates.

USB can support 127 devices in a plug-and-play manner, and they can also be hot-swapped. USB detects when a device is added or withdrawn, and automatically reinitiates the system. USB uses a tiered star topology in which only one device must be plugged into the PC's host (root) connector. There is only one host in the system. It becomes a hub and additional devices can be connected directly to that hub or to additional hubs. The host polls connected devices and initiates all data transfers.

The typical detachable cable is known as an “A to B” cable. “A” plugs are always oriented upstream towards the host and “B” plugs are always oriented downstream toward the USB device.

USB host controllers manage the driver software and bandwidth required by each peripheral connected to the bus and USB hubs can detect attachments and detachments of peripherals using biased termination at cable ends. Hubs are required for multiple connections to the host connectors.

USB provides asynchronous transfer, but isochronous transfer is used for relatively higher bandwidth devices (audio/video).

USB-On-the-Go (USB OTG) supplements the USB specification and is used for portable devices such as cell phones and digital cameras. It allows limited hosting capabilities for direct point-to-point communication with selected peripherals. It is a smaller connection and has low-power features. USB OTG is compliant with the USB 2.0 specification.

Sound Cards

Personal computers can be used to integrate audio, video, animation, graphics, and to author sophisticated multimedia presentations. In order to use these software tools, computer hardware processing, and storage capabilities have been improved, as have computer operating systems.

The MPC (Multimedia Personal Computer) platform specification was introduced by Microsoft as a part of the Multimedia Extensions included in Windows. The MPC standard demands a provision to record and play 8- to 16-bit linear PCM audio, as well as a software-controlled mixer to combine diverse sound sources and channels. Most PC's include digital signal processing, polyphonic synthesis, I/O, and other features. In many cases, these features are contained on sound cards.

Most motherboards or sound cards contain ADC and DAC, as well as hardware and software based processing to permit recording and playback of stereo or 8 or 16-bit audio at multiple sampling rates. They also allow playback via wavetable synthesis, sampled sounds, of FM synthesis. They provide digital I/O for CD or DVD interfaces; software-controlled audio mixer; on board power amplifiers; and some might provide analog line-in and line-out, S/PDIF input and output, microphone input, and a gamepad/joystick MIDI connector.

Some devices allow a DRAM upgrade to augment onboard memory. Sound cards plug in into a ISA or PCI expansion slot and are accompanied by the appropriate software (drivers) that is bundled with the card.

Synthesis capabilities are used to create sound effects when playing MIDI files, and playing video games. Most chip sets support sample-based wavetable synthesis; this allows synthesis and playback of both music and sound effects via software. The chip may support 128 wavetable instruments, or 64 voices if they have multi-timbral capabilities on 16 channels. Other chip sets support physical model synthesis.

Most chips have a MIDI interface for connection to an external MIDI hardware instrument, and some might contain built-in 3-D stereo enhancement circuitry, or dedicated DSP chips.

Music Synthesis

Wavetable synthesis: generate audio signals from a file consisting of a table of audio samples. Traditionally, these tables are filled with single cycles of simple basis waveforms such as sinusoids or triangle waves. Complex sounds are generated by dynamically mixing the simple signals using sophisticated algorithms.

During playback, a pointer loops through the table continuously reading samples and sending them to the D/A converter. Different pitches are obtained by changing the rate at which the table is read. Higher pitches are achieved by skipping samples, while lower frequencies are achieved by adding interpolated samples. A table may be only 512 samples in length, thus generating a low data overhead.

Sample-based synthesis: uses short recordings of musical instruments and other sounds for the basis of the waveforms. These synthesizers may use thousands of samples per table, but only the transient and a small portion of the steady-state are needed for storage. During playback, the transient is read and, if the sound is sustained, the steady-state is looped via a wavetable. This synthesis technique is also known as wavetable synthesis.

Physical modeling synthesis: the physical behavior of an instrument is modeled with software. For example, a plucked string sends transversal waves along the length of the string. As the vibrations move through the string they loose energy and decay. In the software model of a string, an impulse is sent through a circular delay line and it is connected to attenuating filters. The length of the delay line controls the pitch, and filters provide the proper timbre and decay.

Physical modeling is easily implemented in software and can produce consistent results.

Surround Sound Processing

Stereo surround sound expansion programs process stereo signals to enlarge the perceived ambient field. Other 3D positioning programs seek to place sounds in particular locations. Psychoacoustic cues are used to replicate the way sources would sound if they were actually in a 360º space. This processing often uses HRTF's to calculate the sound heard at the listener's ears relative to the spatial coordinates of the sound's origin. These systems process sound during real-time playback, without prior encoding, to position sound statically or dynamically. Although in some cases, the surround process must be encoded in the media itself.

In the consumer audio market, stereo systems have been displaced by multichannel speaker systems (5.1, 7.1). Both Dolby Digital and DTS employ 5.1 channel processing.

These multichannel systems are seen in home theater systems, PC's, and gaming platforms. Although 5.1 playback improves realism, it presents the practical problem of arranging an array of six speakers around a PC. Thus, a number of surround synthesis companies have developed algorithms to specifically replay multichannel formats over two speakers, creating “virtual speakers” to convey the correct spatial sense.

Audio Codec '97 (AC' 97)

The AC '97 component specification describes a two-chip partitioned architecture that provides high-quality PC audio features. Legacy systems integrate the hardware on the ISA bus so that analog circuitry is consolidated with digitally intensive bus interfaces and digital synthesizer circuits, resulting in signal degradation. The AC '97 specification segregates the digital portion of the audio system from the analog portion; it calls for a digital chip (control and DSP) on the bus itself, and an analog chip (interfacing, conversion) off the bus and near the I/O connectors. AC '97 supports all Windows drivers and bus extensions, and it can be used either on the motherboard or on a sound card. It is also backwards compatible with legacy ISA applications.

The specification provides for four analog line-level stereo inputs, two analog line-level monaural inputs, 4- or 6-channel output, I^2S input port, S/PDIF output port, USB and IEEE 1394 ports, and a headphone jack. The specification uses a fixed 48 kHz sampling rate for DVD-Video compatibility with surround tracks coded at 48 kHz.

The AC '97 specification allows for the development of a wide range of chips, with different functions, while retaining their basic compatibility. For example, baseline chips simply connect the computer to a basic analog input/output section. While a more sophisticated chip set might perform digital mixing, filtering, compression/expansion, reverberation, EQ, synthesis, room analysis, and other DSP functions. It can also provide 20-bit conversion, pseudo-balanced analog I/O, and digital interfacing to other protocols. AC '97 can also be used for high-quality stereo playback, 3D audio, multiplayer gaming, and interactive audio/video. AC '97 compliant PC's may contain DVD-ROM drives, TV tuner cards, A/V capture and playback cards, and Dolby Digital decoders.

Windows Multimedia API and DirectX API

An API (Applications Programming Interface) allows developers to access the low-level functions of an operating system so that applications can be built and implemented. In the case of Windows Multimedia API, it allows access to sound card functionality. However, the developer cannot access the peripherals directly and is limited to whatever functions Windows provides. For example, the Multimedia API does no provide any means for mixing audio files.

Microsoft's DirectX API was designed to overcome some of these limitations, and promotes high-performance multimedia application in Windows. DirectX API provides real-time, low-level access to peripherals specifically used in intensive audio/video applications. It also divides multimedia tasks in different components:

DirectSound: provides device-independent access to audio accelerator hardware. Provides functions for mixing audio files and control over each file's volume, balance, and playback rate within the mix. It also allows for low-latency playback.

DirectSound3D API: provides a set of functions that allow application programmers to add 3D audio effects. Processing can be done natively or in an expansion card's hardware.

DirectMusic: provides wavetable synthesis with support for Downloadable Sounds (DLS), interactive music composition, and authoring tools. DLS is an extension of the MIDI specification that defines a file format, device architecture, and API. DLS allows synthesizer developers add custom wavetables to the General MIDI sounds stored in the sound card's ROM.

DirectShow: supports DVD decoders and DVD applications, as well as other aspects such as playback, navigation, regional management, and exchange of CSS encrypted data.

Other components of the DirectX API are DirectDraw, Direct-Play, DirectInput, and DirectSetup.

Vendors usually provide DirectX drivers in addition to their standard Windows drivers. If a DirectX driver is not supplied, DirectX will provide an emulated driver. An emulated driver essentially has the same functionality as an original driver, but it may slower in its performance.

If the audio functionality is implemented in multiple chip designs, usually a sound card was needed. Some manufacturers can consolidate these functions onto a single chip design or “sound card on a chip”. This lowers the cost and allows PC manufacturers to place the audio functions on the motherboard.

MMX

The Intel Multimedia Extensions (MMX) instruction set is contained in a single Pentium processor, and is specially designed to accelerate graphics, video, and audio signal processing. Software written for MMX will run 44-60% faster for certain tasks. This efficiency allows faster execution and frees other system resources for more sophisticated processing. However, software-based processing on the host CPU has limitations. For a processor devotes half of its power to processing an audio application, the remaining half will be devoted to other simultaneous applications.

File Formats

Defined file formats are needed to transfer essence (content data) along with metadata. In this manner, the workflow between different collaborators is greatly improved, as well as allowing the transfer of essence from one platform to another.

Essence or content data can be audio, video, images, graphics, and text. *Metadata* is related data that describes essence, and can include information regarding sampling frequency, downmixing, number of channels, synchronization information, copyright information and ownership.

Audio data can sometimes be stored as raw data or headerless sound file that contains only amplitude samples. Dedicated file formats contain an introductory header with metadata.

Some common file formats for audio and video include WAV, AIFF, SDII, QuickTime, JPEG, MPEG, and OMFI.

Open Media Framework Interchange (OMFI)

OMFI is a set of file format standards for audio, text, still graphics, images, animation, and video files. In addition, it defines editing, mixing, and processing notation so that both content and description of edited audio and video programs can be interchanged. The format contains metadata information as well as sampling and timecode information, and accommodates compressed and uncompressed files. Files can be created in one format, interchanges to another platform for editing and signal processing, then returned to the original format without any information loss. An OMFI file contains all the information needed to create, edit, and play digital media presentations.

OMFI uses two basic types of information: *Compositions* or descriptions of all the data required to play or edit a presentation, and *physical sources* or the actual media data, as well as identification of the sources used in the composition.

OMFI allows file transfer via removable disk exchange and transmission on a fiber network. Common file formats included in the OMFI format are: TIFF, AIFC, and WAV.

Advanced Authoring Format (AAF)

AAF, successor to the OMFI specification, is an open-source interchange protocol for professional multimedia post-production and authoring applications (not delivery); essence and metadata can be exchanged between multiple users and facilities, diverse platforms, systems, and applications. It defines the relationships between content elements, maps elements to a timeline, synchronization between content elements, maps elements to a timeline, synchronizes content to streams, describes processing, track the history to the file, and can reference external essence not in the file.

Material eXchange Format (MXF)

Media format for the exchange of program material used in professional post-production and authoring applications. It provides a simplified standard container format for essence and metadata, and is platform independent. It is closely related to AAF. MXF files can contain any particular media format such as MEG, WAV, AVI, or DPX, and MXF files can be associated with AAF projects. Both MXF and AAF share the same internal data structure, and both use the SMPTE Metadata Dictionary to define data and workflow information.

MXF data structures offers a “partial retrieve” so that users can retrieve only sections of a file that are pertinent to them without copying the entire file. Because essence is placed in a temporal streaming format, data can be delivered in real time, or conveyed with conventional file-transfer operations. MXF data is sequentially arranged and it is commonly used for finished projects and for writing to media.

AES31

AES31 is a group of non-proprietary specifications used to interchange audio data and project information between devices. Simple exchange of one audio file, or exchanges of complex files with editing information from different devices can be accomplished. There are four independent stages with interchange options:

AES31-1: defines the physical data transport describing how files can be moved via removable media or high-speed network.

AES31-2: defines an audio file format, describing how BWF data chunks should be placed on the storage media or packaged for network transfer.

AES31-3: defines a simple project structure using sample-accurate audio decision list (ADL) that contain information such as levels, crossfades, and allows for files to be played back in synchronization. Files use the .adl extension and uses a URL to identify a file source, whether residing locally or on a network. Edit lists are conveyed in human-readable ASCII.

AES31-4: defines an object-oriented project structure capable of describing a variety of characteristics. A universal source locator can access files on different medias or platforms; the locater specifies the file, host, disk volume, directory, subdirectories, and the file name with a .wav extension.

Digital Audio Extraction

Music data is copied from a CD with direct digital means using a CD- or DVD-ROM drive to create a file (WAV) on the host computer. The copy is ostensibly a clone of the original data. The Red Book and Yellow Book formats have different ways in which they store information. The first stores data in tracks, whereas the latter stores the data in files. Thus, when writing to CD-R or CD-RW the data sent to the recorder must be continuous, otherwise if there is an interruption at the writing laser the recording is unusable. Digital audio extraction is also known as “ripping”, which refers to the copying of music into files.

CD- and DVD-ROM drives are capable of digitally transferring information from a CD-ROM disc to a host computer. In addition, they can also play back CD-Audio discs by reading data through audio converters and analog output circuits. Because of the differences between CD-ROM and CD-Audio file formats, digital audio extractors require a special logic (hardware, software, or firmware) on the host. This logic is needed to correct synchronization errors that would result when a CD-Audio is read like a CD-ROm disc.

Hard-Disk Recording

Provides random-access recording on hard drives with back-up on other media; host computer processing of digital audio signals; and programmable output. In addition, the recorder should be highly interactive and promote an intuitive environment for learning its operation. It should provide flexibility and efficiency, as well as fidelity.

Hard-disk recorders consolidate storage, editing, production, and interfacing features. A single recorder takes place of several pieces of traditional studio hardware. Also, production time is decreased and creative possibilities are enhanced thanks to random access editing.

System Features

Hard-disk recorders provide random access, and multitrack recording and playback. System functions include nondestructive editing, DSP for mixing, EQ, compression, reverberation, synchronization to timecode and other timebase references; data back-up; networking; media removability; external machine control; sound-cue assembly; edit decision I/O; and analog and digital data I/O.

In most cases, this can be accomplished with a personal computer, and dedicated audio processing electronics that can interface to it, or software plug-in programs that add specific functionality. Hard-disk recorders provide multitrack operation. Time-division multiplexing is used to overcome limitations of hard-wired bus structure, so that the number of tracks does not equal the number of audio outputs; software systems are much more flexible than hard-wired systems.

Hard-disk recorders use a graphical interface, with most human actions taking place with a mouse and keyboard; some systems provide a dedicated hardware controller that can remotely control the system.

Most systems provide a standard “tape recorder” transport control section that features autolocation points, punch-in/-out indicators, time scale indicators. Grabber and zoomer tools, as well as fading, crossfading, gain change, tempo and pitch change, etc.

Hard-disk recorders can provide virtual mixing capabilities of both audio and MIDI tracks. Nondestructive bouncing allows tracks to be combined during production, prior to mixdown.

DSP hardware and software perform specific “effects” on the programs loaded onto memory.

Although most hard-disk recorders use Mac or PC computers, some are stand-alone products designed with a user interface that emulate traditional multitrack tape recorders. The dedicated system provides random access, audio scratchpad, editing, punch-in, crossfading, and set cue points and loops. Finished material can be exported to different media, such as optical media.

Time-division multiplexing (TDM) provides a means of connection and routing between devices. The TDM bus provides a software-controlled routing matrix with many sources and destinations including software modules, and analog and digital external devices. The bus allows routing of 256 channels of 24-bit data at variable sample rates.

The primary audio storage is handled remotely with hard-disk and optical discs. The drives can be interfaces to the workstation via SCSI ports. Multiple SCSI controllers are needed to connect and entire bank of drives.

Peripheral devices, or other systems, need to interface to the output storage medium. High-quality I/O is critical under these circumstances. Transmission protocols, such as AES3, S/PDIF, and IEEE 1394, can link the elements in an all digital digital studio

Flash Memory

Portable flash memory offers a small and robust way to store data to nonvolatile solid-state memory (electrically erasable programmable read-only memory or EEPROM), via an onboard controller. There are several flash memory formats with a different form factor, such as Compact Flash, Memory Stick, SD, and others. Flash memory cards are designed to directly plug into a receiving socket, but other devices interface via USB or other means.

Flash memory technology allows high-storage capacity and high data transfer rates. Some cards also incorporate WiFi or Bluetooth for wireless data exchange and communication.

Desktop Audio Software Applications

The low cost of PC's and software applications has encouraged their wide use by musicians and home recordists. Both software and hardware quality is remarkable. Numerous PC software packages permit audio recording, editing, processing, and analysis. The PC acts as a hard-disk digital recorder for CD-compatible audio quality, or other sampling and word lengths. Different software packages permit operation in different time modes: samples, absolute frames, measures and beats, and different SMPTE timecodes. They also perform audio file conversion and support different audio file types. MIDI interfacing is also possible.

What can a desktop audio program do?

...

What are some of the applications?

...

Workstations can combine different elements of a recording studio with those of in video production studio.

Hard-disk Editing

Audio editing is greatly facilitated with random access storage, instantaneous auditioning, level adjusting, marking, crossfading, and nondestructive editing. Editing errors can be immediately corrected with an undo command.

Using an edit cursor, clipboard, cut-paste, and other tools, sample accurate editing is easily accomplished. Edit points are located in ways analogous to analog tape recorders; sound is “scrubbed” back and forth until the edit point is found (by reading from memory). In some cases, edit points are assigned by entering timecode number. Crossfades can be done automatically or manually.

What an audio editor do?

…

Audio for Video Workstations

Some audio workstations implement digital video displays for video editing, and for synchronizing audio to picture. Users can prepare audio tracks for QuickTime movies, with random access to many takes. Using authoring tools, audio and video materials can be combined into a final presentation, such as DVD-ROM.

Video edit systems can be classified as follows:

Nonlinear systems: disk-based.

Linear systems: videotape-based.

Off-line systems: edit audio and video programs, generate an edit decision list (EDL)

On-line systems: higher-quality video system used for the final assembly of materials.

Audio for video workstations offer a selection of frame rates and sampling frequencies, as well as other video prerequisites. Some also provide direct control over external videotape recorders.

Depending on the application, either audio or video can be designated as the master providing different functionalities as required by the user.

During a video session, both video and audio can be captured via analog inputs. There are different options for digitizing audio and video, as well as file format. Editing of both audio and video, as well as linking them, can be done in authoring software. Finally, finished projects can be compiled into a destination file, such as QuickTime movie, or can be recorded to an external videotape or optical recorder.