UNIT 2 Sound/Audio

Objectives

- To understand how computers process sound
- To understand how computers synthesize sound
- To understand the differences between two major kinds of audio, namely digitised sound and MIDI music

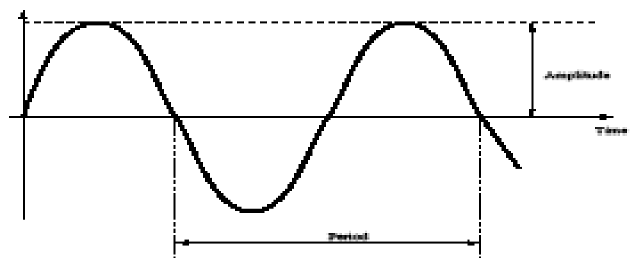
Contents

1 The Nature of Sound

Sound is a physical phenomenon produced by the vibration of matter and transmitted as waves. However, the perception of sound by human beings is a very complex process. It involves three systems:

- the *source* which emits sound;
- the *medium* through which the sound propagates;
- the *detector* which receives and interprets the sound.

Sounds we heard everyday are very complex. Every sound is comprised of waves of many different frequencies and shapes. But the simplest sound we can hear is a sine wave.



Sound waves can be characterised by the following attributes:

Period	Frequency	Amplitude	Bandwidth
	Pitch	Loudness	Dynamic

1.1 Pitch and Frequency

Period is the interval at which a periodic signal repeats regularly.
Pitch is a perception of sound by human beings It measures how 'high' is the sound as it is perceived by a listener.
Frequency measures a physical property of a wave. It is the reciprocal value of period $f = \frac{1}{P}$. The unit is Herts (Hz) or kiloHertz (kHz).

Musical instruments are tuned to produce a set of fixed pitches.

Infra-sound	0 – 20 Hz
Human hearing range	20 – 20 kHz
Ultrasound	20 kHz – 1 GHz
Hypersound	1 GHz – 10 THz

Note	Ratio	Frequencies
C	1:1	264
D	9:8	297
E	5:4	330
F	4:3	352
G	3:2	396
A	5:3	**440**
B	15:8	495
C	2:1	528

1.2 Loudness and Amplitude

The other important perceptual quality is *loudness* or *volume*.

Amplitude is the measure of sound levels. For a digital sound, amplitude is the sample value.

The reason that sounds have different loudness is that they carry different amount of power. The unit of power is watt. The intensity of sound is the amount of power transmitted through an area of $1m^2$ oriented perpendicular to the propagation direction of the sound.

If the intensity of a sound is $1watt/m^2$, we may start feel the sound. The ear may be damaged. This is known as the *threshold of feeling*. If the intensity is $10^{-12}watt/m^2$, we may just be able to hear it. This is know as the *threshold of hearing*.

The relative intensity of two different sounds is measured using the unit *Bel* or more commonly *deciBel* (*dB*). It is defined by

$$\text{relative intensity in } dB = 10 \log \frac{I_2}{I_1}$$

Very often, we will compare a sound with the *threshold of hearing*.

160 dB	Jet engine
130 dB	Large orchestra at fortissimo
100 dB	Car on highway
70 dB	Voice conversation
50 dB	Quiet residential areas
30 dB	Very soft whisper
20 dB	Sound studio

Typical sound levels generated by various sources

Intensity ($watt/m^2$)	**Sound Level** dB	***Loudness***
1	120	*Threshold of feeling*
10^{-3}	90	*fff*
10^{-4}	80	*ff*
10^{-5}	70	*f*
10^{-6}	60	*mf*
10^{-7}	50	*p*
10^{-8}	40	*pp*
10^{-9}	30	*ppp*
10^{-12}	0	*Threshold of hearing*

Typical sound levels in music

1.3 Dynamic and Bandwidth

- *Dynamic range* means the change in sound levels.

  For example, a large orchestra can reach 130dB at its climax and drop to as low as 30dB at its softest, giving a range of 100dB.

- *Bandwidth* is the range of frequencies a device can produce or a human can hear.

FM radio	50Hz – 15kHz
AM radio	80Hz – 5kHz
CD player	20Hz – 20kHz
Sound Blaster 16 sound card	30Hz – 20kHz
Inexpensive microphone	80Hz – 12kHz
Telephone	300Hz – 3kHz
Children's ears	20Hz – 20kHz
Older ears	50Hz – 10kHz
Male voice	120Hz – 7kHz
Female voice	200Hz – 9kHz

2 Computer Representation of Sound

- Sound waves are continuous while computers are good at handling discrete numbers.
- In order to store a sound wave in a computer, samples of the wave are taken.
- Each sample is represented by a number, the 'code'.
- This process is known as *digitisation*.
- This method of digitising sound is know as *pulse code modulation* (PCM).

  Refer to Unit 1 for more information on digitisation.
- According to Nyquist sampling theorem, in order to capture all audible frequency components of a sound, i.e., up to $20kHz$, we need to set the sampling to at least twice of this.

  This is why one of the most popular sampling rate for high quality sound is $4410Hz$.
- Another aspect we need to consider is the resolution, i.e., the number of bits used to represent a sample.

  Often, 16 bits are used for each sample in high quality sound. This gives the SNR of $96dB$.

2.1 Quality versus File Size

The size of a digital recording depends on the sampling rate, resolution and number of channels.

$$S = R \times (b/8) \times C \times D$$

Higher sampling rate, higher resolution gives higher quality but bigger file size.

S	file size	bytes
R	sampling rate	samples per second
b	resolution	bits
C	channels	1 - mono, 2 - stereo
D	recording duration	seconds

For example, if we record 10 seconds of stereo music at 44.1kHz, 16 bits, the size will be:

$$\begin{aligned} S &= 44100 \times (16/8) \times 2 \times 10 \\ &= 1,764,000\text{bytes} \\ &= 1722.7\text{Kbytes} \\ &= 1.68\text{Mbytes} \end{aligned}$$

Note: $\begin{aligned} 1\text{Kbytes} &= 1024\text{bytes} \\ 1\text{Mbytes} &= 1024\text{Kbytes} \end{aligned}$

High quality sound files are very big, however, the file size can be reduced by compression.

File size for some common sampling rates and resolutions

Sampling Rate	Resolution	Stereo /Mono	Size for for 1 Min.	Comments
44.1KHz	16-bit	Stereo	10.5MB	CD-quality recording
44.1KHz	16-bit	Mono	5.25MB	A good trade-off for high-quality recordings of mono sources such as voice-overs
44.1KHz	8-bit	Stereo	5.25MB	Achieves highest playback quality on low-end devices such as most of the sound cards
44.1KHz	8-bit	Mono	2.6MB	An appropriate trade-off for recording a mono source
22.05KHz	16-bit	Stereo	5.25MB	Darker sounding than CD-quality recording because of the lower sampling rate
22.05KHz	16-bit	Mono	2.5MB	Not a bad choice for speech, but better to trade some fidelity for a lot of disk space by dropping down to 8-bit
22.05KHz	8-bit	Stereo	2.6MB	A very popular choice for reasonable stereo recording where full bandwidth playback is not possible
22.05KHz	8-bit	Mono	1.3MB	A thinner sound than the choice just above, but very usable
11KHz	8-bit	Stereo	1.3MB	At this low a sampling rate, there are few advantages to using stereo
11KHz	8-bit	Mono	650K	In practice, probably as low as you can go and still get usable results
5.5KHz	8-bit	Stereo	650K	Stereo not effective
5.5KHz	8-bit	Mono	325K	About as good as a bad telephone connection

2.2 Audio File Formats

The most commonly used digital sound format in Windows systems is `.wav` files.

- Sound is stored in `.wav` as digital samples known as *Pulse Code Modulation*(PCM).
- Each `.wav` file has a header containing information of the file.
  - type of format, e.g., PCM or other modulations
  - size of the data
  - number of channels
  - samples per second
  - bytes per sample
- There is usually no compression in `.wav` files.

Other format may use different compression technique to reduce file size.

- `.vox` use *Adaptive Delta Pulse Code Modulation* (ADPCM).
- `.mp3` MPEG-1 layer 3 audio.
- RealAudio file is a proprietary format.

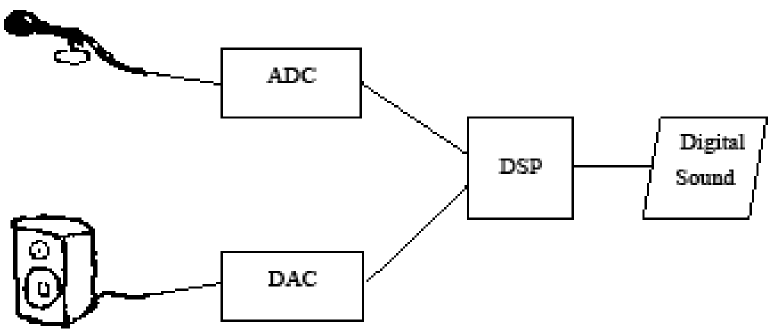
Some common audio files formats

Extension	**MIME Type**	Platform	Use
aif	Audio/x-aiff	Mac, SGI	Audio
aifc	Audio/x-aiff	Mac, SGI	Audio (compressed)
AIFF	Audio/x-aiff	Mac, SGI	Audio
aiff	Audio/x-aiff	Mac, SGI	Audio
au	Audio/basic	Sun, NeXT	ULAW audio data
mov	Video/QuickTime	Mac, Win	QuickTime video
mpe	Video/mpeg	All	MPEG video
mpeg	Video/mpeg	All	MPEG video
mpg	Video/mpeg	All	MPEG video
mp3	Audio/x-mpeg	All	MPEG audio
qt	Video/QuickTime	Mac, Win	QuickTime video
ra,ram	Audio/x-pn-realaudio	All	RealAudio Sound
snd	Audio/basic	Sun, NeXT	ULAW Audio Data
vox	Audio/	All	VoxWare Voice
wav	Audio/x-wav	Win	WAV Audio

2.3 Audio Hardware

- Recording and Digitising sound:
  - An *analog-to-digital converter*(ADC) converts the analog sound signal into digital samples.
  - A *digital signal processor*(DSP) processes the sample, e.g. filtering, modulation, compression, and so on.
- Play back sound:
  - A *digital signal processor* processes the sample, e.g. decompression, demodulation, and so on
  - An *digital-to-analog converter*(DAC) converts the digital samples into sound signal
- All these hardware devices are integrated into a few chips on a sound card

- Different sound card have different capability of processing digital sounds. When buying a sound card, you should look at:
  - maximum sampling rate
  - stereo or mono
  - duplex or simplex

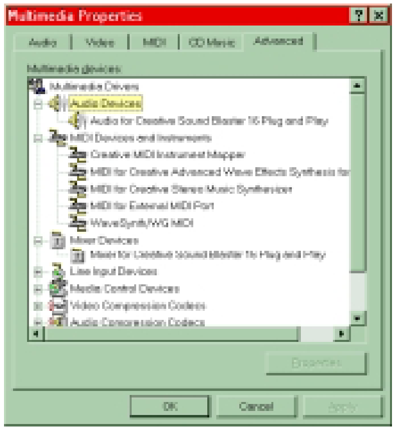


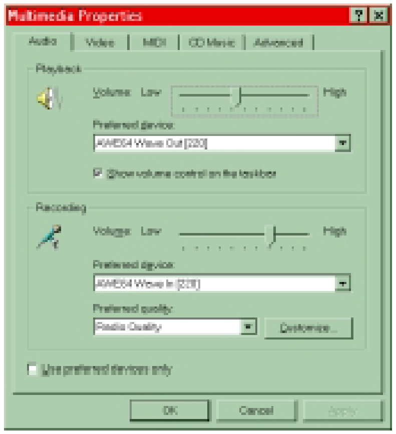
2.4 Audio Software

- Windows device driver — controls the hardware device.

  Many popular sound cards are Plus and Play. Windows has drivers for them and can recognise them automatically. For cards that Windows does not have drivers, you need to get the driver from the manufacturer and install it with the card.

- If you do not hear sound, you should check the settings, such as interrupt, DMA channels, and so on.





- Device manager — the user interface to the hardware for configuring the devices.
  - You can choose which audio device you want to use
  - You can set the audio volume

Mixer — its functions are:

- to combine sound from different sources
- to adjust the play back volume of sound sources
- to adjust the recording volume of sound sources

Recording — Windows has a simple Sound Recorder.

Editing — The Windows Sound Recorder has a limiting editing function, such as changing volume and speed, deleting part of the sound.

There are many freeware and shareware programs for sound recording, editing and processing.

3 Computer Music — MIDI

Sound waves, whether occurred natural or man-made, are often very complex, i.e., they consist of many frequencies. Digital sound is relatively straight forward to record complex sound. However, it is quite difficult to generate (or synthesize) complex sound.

There is a better way to generate high quality music. This is known as *MIDI* — Musical Instrument Digital Interface.

It is a communication standard developed in the early 1980s for electronic instruments and computers. It specifies the hardware connection between equipments as well as the format in which the data are transfered between the equipments.

Common MIDI devices include electronic music synthesisers, modules, and MIDI devices in common sound cards.

General MIDI is a standard specified by MIDI Manufacturers Association. To be GM compatible, a sound generating device must meet the General MIDI system level 1 performance requirement.

- minimum of 24 fully voices
- 16 channels, percussion on channel 10
- minimum 16 simultaneous and different timbre instruments
- minimum 128 preset instruments
- Support certain controllers

This sign indicated that the device is a general MIDI device.

GENERAL MIDI

3.1 MIDI Hardware

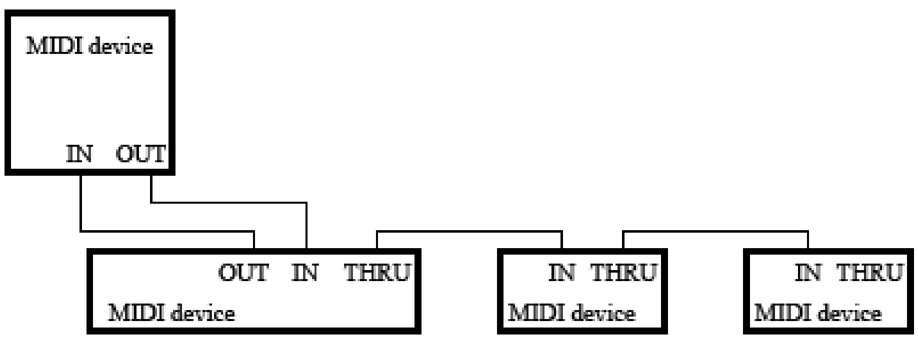
An electronic musical instrument or a computer which has MIDI interface should has one or more MIDI ports. The MIDI ports on musical instruments are usually labelled with:

IN — for receiving MIDI data;

OUT — for outputting MIDI data that are generated by the instrument;

THRU — for passing MIDI data to the next instrument.

MIDI devices can be daisy-chained together.



MIDI software

MIDI player for playing MIDI music. This includes:

- Windows media player can play MIDI files
- Player come with sound card — Creative Midi player
- Freeware and shareware players and plug-ins— Midigate, Yamaha Midplug, etc.

MIDI sequencer for recording, editing and playing MIDI

- Cakewalk Express, Home Studio, Professional
- Cubasis
- Encore
- Voyetra MIDI Orchestrator Plus

Configuration — Like audio devices, MIDI devices require a driver. Select and configure MIDI devices from the control panel.

4 Summary — MIDI versus digital audio

Digital Audio	MIDI
• Digital representation of physical sound waves	• Abstract representation of musical sounds and sound effects
• File size is large if without compression	• MIDI files are much more compact
• Quality is in proportion to file size	• File size is independent to the quality
• More software available	• Much better sound if the sound source is of high quality
• Play back quality less dependent on the sound sources	• Need some music theory
• Can record and play back any sound including speech	• Cannot generate speech