FACTFILE:
GCE DIGITAL TECHNOLOGY

AS2: FUNDAMENTALS OF DIGITAL TECHNOLOGY

HARDWARE AND SOFTWARE - DATA COMPRESSION

Learning Outcomes

Students should be able to:
- explain the need for data compression;
- describe how zipping is used to compress data;
- evaluate common data file formats: txt, wav, bitmap, JPEG, MPEG and GIF;
- describe how text, sound and video can be input using a range of devices, such as personal computers, laptops, tablets or smartphones.

Content in Data Compression

- Introduction – Why is data compression necessary?
- How zipping is used to compress data
- Evaluating data file formats
  - txt
  - wav
  - bitmap
  - JPEG
  - MPEG
  - GIF
- Inputting text, sound and video using digital devices
- Questions

DATA COMPRESSION: INTRODUCTION

Why is data compression necessary?
Data compression is the process associated with reducing the memory or storage required by large files. This is often important when it comes to data transmission and storage on backing storage devices. Data compression uses a series of algorithms to reduce the amount of real space taken up by data on a storage medium.

How zipping is used to compress data
Zipping is the process of compressing data files. It uses an algorithm called the LZW algorithm to reduce the memory or storage required by large files. The LZW algorithm looks for repeating patterns in the data being compressed and will then replace these repeating patterns with a single character.

For example in the first two sets of paragraphs and titles in this fact file there are 10 instances of the 'ss', if these were all replaced by * this would save 10 characters.

Can you see any other repetitions where patterns /pairs of characters could be replaced by a single character, thus reducing the text file size even further?

Evaluating Data File Formats
All files written to a storage medium must have a unique file name. The first part of the file name

will be determined by the user and will help them identify the file at a later date; the second part (the file extension) will help identify that file as being a certain file type. Some of the more common file extensions are referenced below.

txt – files which are presented as lines of electronic text. These files are readable to the human eye but contain no formatting information and may often be used to store information which requires further processing by another application. Some operating systems will place an EOF (End of file character) after the last line in the text file to denote the end of the txt file.

wav – a file format widely used for professional recording and editing. These files use a process known as sampling to store a digital representation of a recorded analogue sound signal. During the sampling process the amplitude of the waveform being received are analysed very quickly and recorded (see diagram below). It is the angular waveform shown below that is played back to the listener. The quality of the sound is dependent upon the frequency of the sampling (sampling rate) and the number of bits used to store the digital value for each sample (sample resolution).
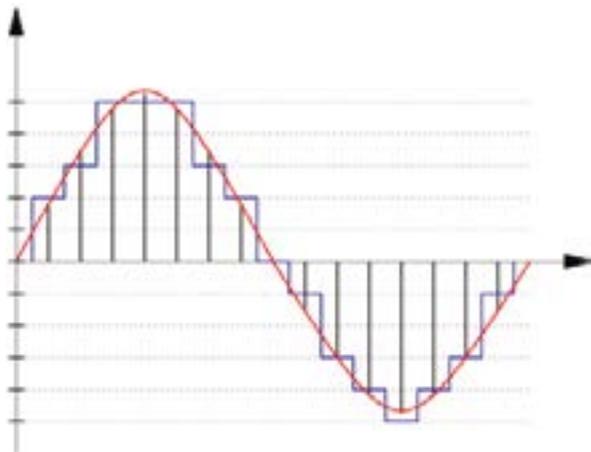
Wave Form Audio File Format (WAVE) but commonly known as WAV due to its filname extension.

Generally a WAV file is uncompressed, although wav files can sometimes also be used to store compressed formats. A header in the file will indicate if the file is in compressed or uncompressed format.

bmp – a method of creating images where details of each pixel forming part of the overall image are held as a bit map in memory. A bit map is where a pattern of bits is used to hold data relating to the state of an individual pixel in the image; including text. Since a bitmap image is produced from tiny squares of colour, which are arranged to produce the effect of an image; this is a good method of reproducing 'continuous tone' images, such as photographs and for free hand drawings.

jpeg – jpegs identify arrangements of pixels which are repeated elsewhere in the image so this data needs only be saved once, rather than having to repeat the data and take up storage space unnecessarily.

mpeg – mpegs use a method called delta compression to record and transmit the data representing audio and video files. It works by only sending what has changed since the last recording / transmission of data. For example in the transmission of TV signals where 25 frames are transmitted per second, a full frame / picture is only sent occasionally and in between those transmissions data is sent relating only to changes in the full frame / picture.

gif – a method developed to support the compression and storage of images using bit mapped data, a simple animated version is also available. Eight bits are used to represent data relating to each pixel so only 256 distinct colours can be represented; helping to minimize file size.

Data File Format	File Type	Advantages	Disadvantages
txt	.txt A text file based commonly on the ASCII character set.	Can be opened in any word processor or text editor without the need for additional processing. Is much smaller than a word processing file. Simple txt files do not require metadata so therefore if no data is stored the file will take up zero bytes. When data corruption occurs on a txt file, the remaining contents can be more easily recovered	Files of this type have no formatting information. These file types can often take up more storage than is necessary as no compression occurs.
wav	.wav Waveform audio file format. A sound file.	A 'lossless' method of recoding sound files, i.e. all of the information collected in the original sampling process will be retained.	Sound quality is dependent upon sample rates and sample resolution.
Bitmap	.bmp	Used to store characters or graphics as individual pixels; the higher the resolution of the image the better the quality of the image. High resolution bit mapped images are ideal for storing freehand drawings	'Blank' portions of the image still require representation within the bit map so this can contribute to the file size of the stored image. Bit mapped images may appear jagged (pixelated) when scaled up or down.
JPEG (defined by the Joint Photographic Experts Group)	.jpeg, .jpg, .jpe A graphics file format, used to store still images	Standard compression format for photographic images devised by the Joint Photographic Experts Group. Can reduce file size to 5% of its original size.	A 'lossy' compression method, i.e. some detail relating to the image content is lost in the compression process.
MPEG (defined by the Motion Picture Experts Group)	MPEG – 1 (mpg) MPEG-2 (m2v, mpg, mp2) Various audio and video formats.	Varying resolutions can be applied to the file at the time of storage for example:– MPEG-1 file formats produce low-resolution file formats suitable for CD storage. MPEG-2 file formats record high-resolutions sequences suitable for DVD recording. Other variations / resolutions are available.	'Lossy' compression techniques are applied i.e. some data is removed. But the degeneration of the image cannot always be perceived by the human eye.
GIF (Graphics Interchange Format)	.gif	Provides a format for image files that supports both animated and static images. Gif's are compressed to save transmission time.	Colour palette is limited to 256 colour's so they are not ideal for storing digital photos, such as those captured with a digital camera.

Inputting text, sound and video using digital devices

In this section we will look at how text, sound and video can be input a collection of digital devices including personal computers, laptops, tablets and smartphones.

Text input – text may be input to digital devices using a variety of methods; the most common being via a keyboard. The QWERTY keyboard is the most common keyboard format available on digital devices; it will incorporate letters, digits and a range of special characters. Some devices may however have a separate numeric keypad.

The set of characters represented on a digital device at any one time is known as the character set of that device. ASCII (American Standard Code for Information Interchange) is a common character set on digital devices. The ASCII code uses 7 bits which gives 32 control codes and 96 displayable characters or symbols. The eighth bit can be used for error checking.

The depression / selection of a key on any device will generate a digital signal representing the character code. When a symbol representing a particular character is to be printed or displayed on the screen of a particular device the character code is converted into the appropriate symbol

Sound input – most digital devices are now supplied with simple microphones for recording sound. Since sound is produced by the vibration of air it can be said to be analogue in format (i.e. it is a continually varying signal). The electrical output from a microphone also consists of a continually varying signal in the form of voltage. Devices such as personal computers, laptops, tablets or smartphones all store data in digital format so conversion of the analogue signal and the digital data is needed. This is achieved through the process of sound sampling where special hardware measures the level of the sound many times and then records this as a numerical value *(see the previous section on wav file formats for additional detail on sampling).*

Video input – it is common place now for digital devices to have a camera incorporated for the recording of digital still and digital video images. Digital video cameras have two main components; the camera, which is the image capturing component, and the recorder which is the component of the video camera that stores the captured video images. The image is recorded using the digital camera. The captured image is then transmitted into computer pixels which is then stored in whatever media storage is available to the digital video camera.

? Questions

1 An image can be stored on a PC using bit-mapped and jpeg file format.
   a. Identify two characteristics of bit-mapped images.   [2]
   b. i. Describe how file size of the same image stored in jpeg format might differ from its
          bit-mapped version.   [1]
      ii. Give one reason for this difference in file size   [2]

....

....

....

....

....

....

2 Sound sampling is used when sound is being recorded on a digital device.
   a. Explain how the process of sampling is used to convert an analogue sound wave
      into digital format for storage on a PC.   [3]
   b. Describe two factors which may impact upon the quality of a digital sound file as
      it is being recorded.   [4]

....

....

....

....

....

....

3 A professional photographer is importing digital images to a PC for detailed editing.
   a. Which image file type would be most appropriate for this task? [1]
   b. Give an explanation for your choice of file type in this instance. [2]

....

....

....

....

....

....

....

4 A student uses a microphone to record a voice over for a multimedia presentation. Describe how analogue sound input is converted into digital format.

....

....

....

....

....

....

....

5 Keyboards are commonly used to enter text into digital devices. Describe how a keyboard can be used to generate text, numbers and special characters which can be stored in digital format. [2]

....

....

....

....

....

....

....

Bibliography
BCS Academy Glossary Working Party, 2013, *BCS Glossary of Computing and ICT,* 13th Edition, Swindon, BCS Learning and Development Ltd