FACTFILE:
GCSE
DIGITAL TECHNOLOGY
Unit 1 – DIGITAL DATA

Fact File 4: Portability

Learning Outcomes

Students should be able to:

- demonstrate understanding of data portability and the following file formats that support it: jpeg, tiff, png, pict, gif, txt, csv, rtf, mp3, mp4, midi, mpeg, avi, pdf, wav and wma.
- demonstrate understanding of the need for data compression.

The Need For Data Compression

Compression means reducing the storage requirements of a file by following one or more compression algorithms.

LOSSLESS compression enables the original file to be fully restored to its original state. LOSSY compression involves discarding some data as part of the overall compression process in such a way that the data can never be reinstated.

Files that demand high volumes of storage also require high bandwidth communication links in order for transmission to be completed within an acceptable time frame. Compression is used to reduce the:

- VOLUME of storage required on storage DEVICES, enabling larger quantities of data to fit into the same amount of storage space
- BANDWIDTH required to transfer multimedia files across COMMUNICATION LINKS (locally, nationally and globally)

Without compression:

- Very high VOLUMES of storage would be required, occupying more physical space and slowing down retrieval/save times
- The available BANDWIDTH on most local and global networks would not cope with the volumes of data being sent, disrupting communication services (leading to e.g. lagging video streaming)

Of course, there will be a price to pay for the benefits of compression, i.e. time will have to be set aside to compress and decompress the files as they are moved in and out of storage or across communication links. If storage/bandwidth is a scarce resource and compression can be done in a tolerable time frame, only then does it make sense.

Data Portability

Portable data is data that can be directly transferred between hardware devices or programs without having to be entered again.

Data needs to be stored using different formats to ensure data is portable.

Image File Types

JPEG

JPEG = Joint Photographic Experts Group

JPEG is a compressed, bitmapped image file type that keeps file sizes low by reducing repetition of pixel patterns throughout the image. It was originally designed for photographs. Nowadays it is a widely used file type for web pages thanks to the effectiveness with which it compresses images. The amount of compression can be varied, and at high levels of compression a very low file size can be achieved, albeit at the expense of the quality of the image. JPEG does not support animation or transparency, and the compression used is lossy which means that image details get permanently lost during compression.

GIF

GIF = Graphics Interchange Format

GIF is a compressed, bitmapped image file type. GIF files can be animated sequences of images or singular still images. A GIF image can contain transparent pixels. Lossless compression techniques are employed when GIF files are saved. Their file sizes are kept low by restricting the colour depth to just 256 colours (which is 8-bit colour, yielding 28 colours).

PNG

PNG = Portable Network Graphics

PNG (pronounced "ping") is a compressed, bitmapped image file type that was initially developed for online image sharing. PNG files do not therefore use the CMYK colour mode that is geared towards printer ink colours. Like GIF files they employ lossless compression techniques and support transparency, but they do not support animation.

TIFF

TIFF = Tagged Image File Format

TIFF images were developed to be usable in both Microsoft Windows and Apple Macintosh operating systems. They are used as a standard file type in commercial printing and OCR software commonly produces TIFF images from scans. TIFF files can behave like containers for holding images of other file types. Their structure is highly complex compared to other image file types and they can use both lossy and lossless compression techniques.

PICT

Apple developed PICT as their native graphic file type for the interchange of graphics between various Apple programs, but in recent years support has reduced and the format has been superseded by PDF since the release of Mac OS X. PICT is an object-oriented file type that can store bitmap and vector graphics. It can use up to 24-bit colour, yielding 224 or 16,777,216 colours which is known as true colour.

Text File Types

TXT

Plain text files can be stored as .TXT (pronounced "text") files. The underlying encoding may vary (e.g. ASCII or ISO-8859 may be used as the character set), but ultimately a .TXT file will store character strings that can be readily viewed and edited by the user in many word processors and other software that is capable of handling plain text (such as code editors and web authoring software). No additional formatting information will be associated with the text itself, so whichever text editing program is used to open the file, its entire contents will be presented textually and not interpreted as formatting information.

CSV

CSV = Comma Separated Values

Data that is structured in a tabular manner will have a common structure on every row. When it is stored in a spreadsheet or database, the tabular structure has clearly visible columns. In the CSV file format, data values are organised into rows that get stored on separate lines, but instead of being spaced out into columns, a comma character (or suitable alternative) is used to separate the data values into their respective columns. CSV files are thus plain text files that can be created, viewed and edited in a plain text editor, but they can equally be opened and edited using tools within spreadsheet and database packages which will interpret the comma as a column delimiter and the new lines as separators between rows. The example below shows a spreadsheet called ClubRaffle being created in spreadsheet software and later saved in the CSV format. The emboldening of the headings is not stored during conversion to CSV, as is the right and left alignment of the data values in the cells of the spreadsheet.


RTF

RTF = Rich Text Format

Formatting information stored amongst plain text is known as markup. Markup is interpreted by word processors and used to show the text in the colours and styles indicated in the markup. However, different word processors use various different ways of encoding their markup, reducing interoperability and preventing some word processors from being able to correctly display files of certain other text file types.

Rich Text Format (RTF) files can translate some of a word processor's own markup into a standardised RTF markup. RTF therefore enables text with some key formatting to be saved and exchanged between different word processors. However, not all of the advanced formatting made possible by some powerful word processors will be retained when it is saved as RTF. Formatting tools found in RTF files include:

• Emboldening, italicising and underlining;
• Left, centre, right and full justification;
• Font faces, colours and sizes;
• Indentation and line spacing adjustment;
• Highlighting;
• Bullets and numbering;
• Image insertion.
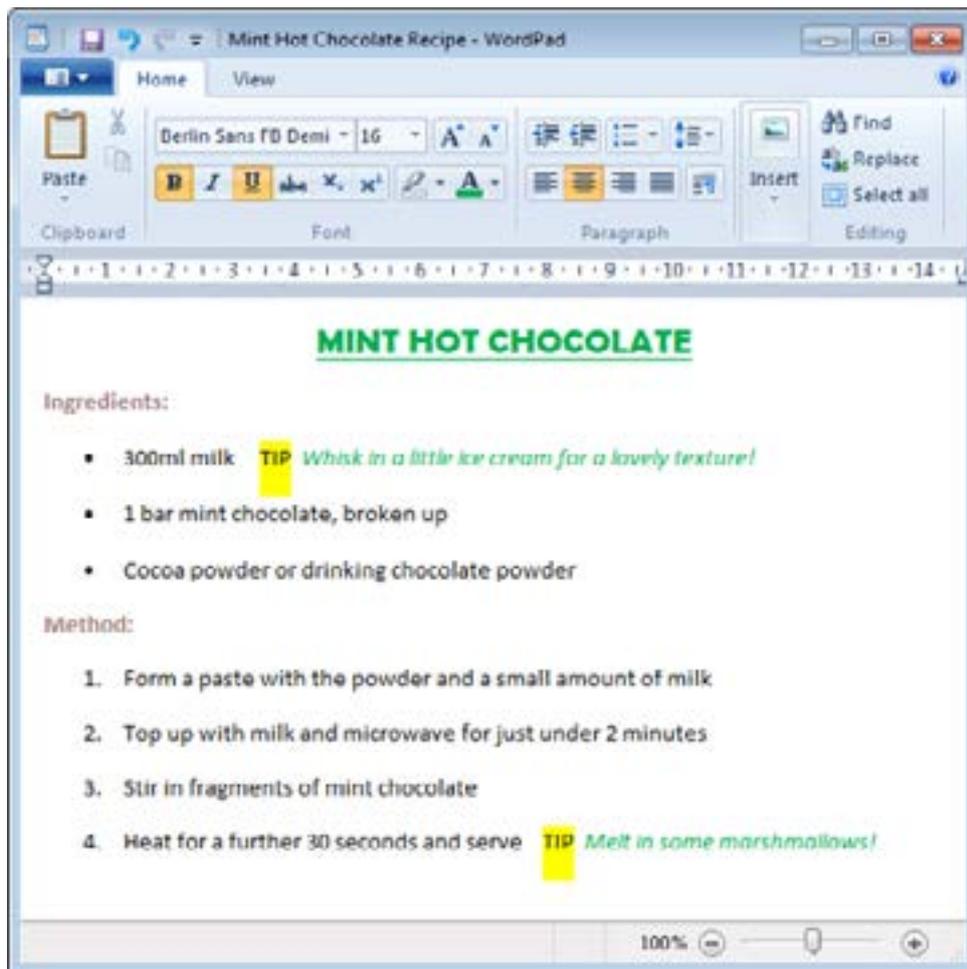
This recipe is stored in Rich Text Format.

PDF

PDF = Portable Document Format

A PDF file holds text, formatting, images and potentially some editable regions which can be filled in by the user. The file type can be viewed using many applications across a wide range of operating systems and hardware platforms, making it very desirable for uploading documents for reading online. PDF reader software can either be downloaded for free or it can be found embedded into commonly-used programs such as web browsers.

Creating a PDF can be achieved using a save or print command from within a program or via a specialised converter tool. Graphics, spreadsheets, web pages and several other files can be exported to PDF format.

Advanced features of PDF files include the following:

• Powerful assistive technology embedded for those who are visually impaired;

• Powerful compression techniques to keep the file size very low;

• Password protection (and the use of digital signatures) to reduce the likelihood of breaches of copyright or to restrict access;

• PDF files can store characters of specific fonts within the document so that there is no substitution of the fonts of the document when it is opened on a computer on which the fonts used are not installed;

• The use of metadata for improving the speed of text searches and for labelling the document with its author, date of creation, document title and so on.

Sound File Types

A waveform is a pattern that represents a sound, i.e. a pattern of vibrations. To convert a sound wave into a digital representation for use on a computer, the wave amplitude is measured at regular time intervals, and this process is called sampling. The number of samples per second (the sample rate) is a major factor affecting the quality of the digital version of the sound file.

Below we see an analogue sound wave and its corresponding digital representation, i.e. a sequence of amplitude measurements taken at regular intervals, with the intermediate values being lost.
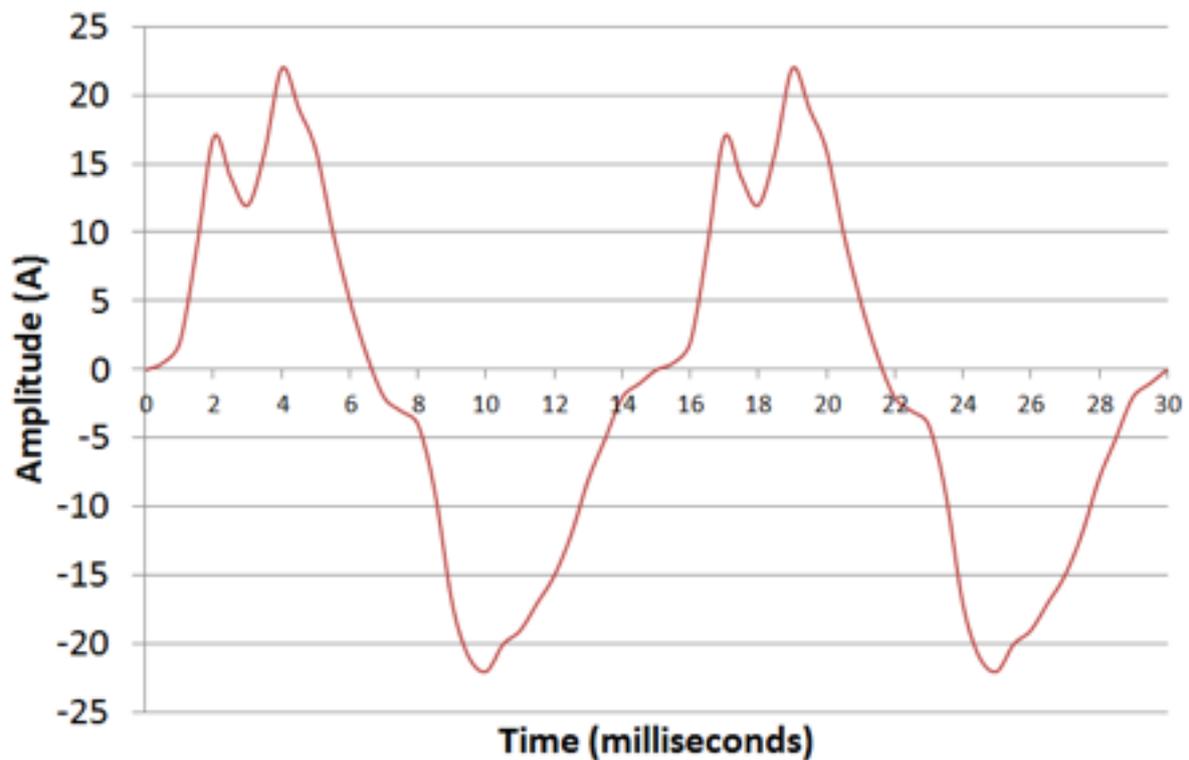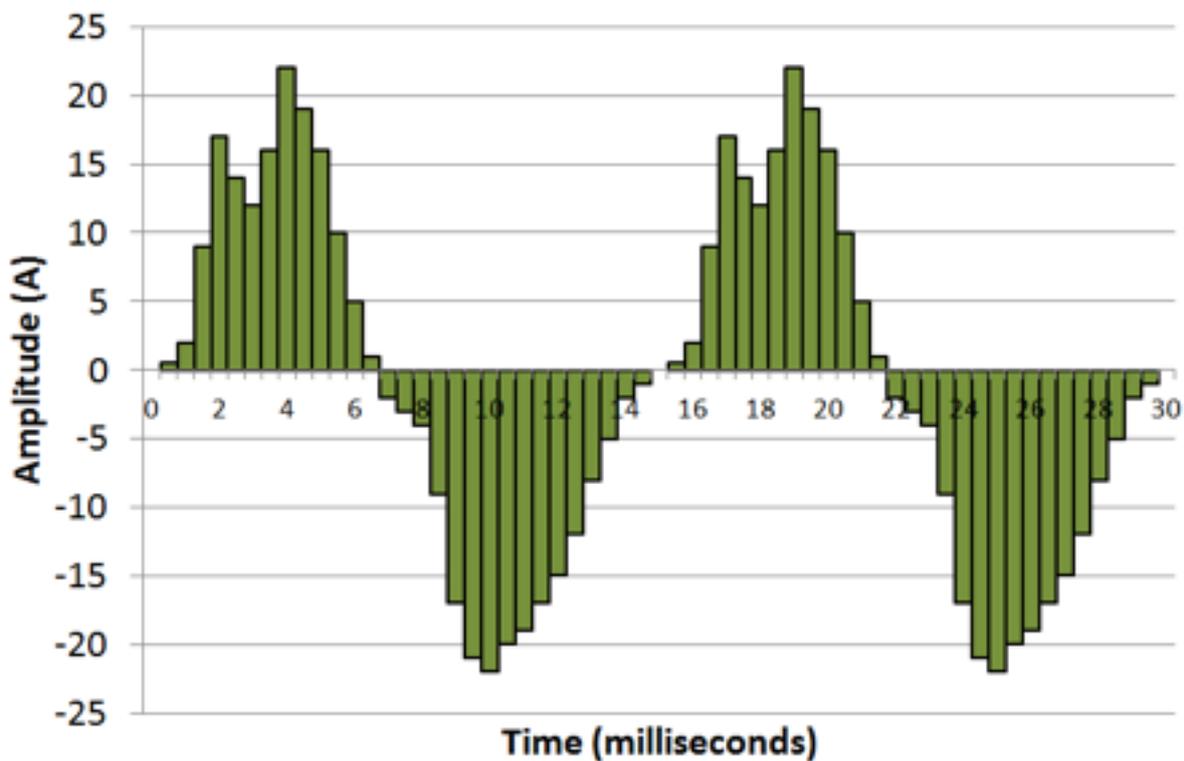
Whilst all sampled sound file types are captured in this way, there are differences in how the sound is subsequently saved.

WAV

WAV = Windows Waveform

The WAV (pronounced "wave") file format is lossless and thus provides a close likeness to the original analogue sound. A wide range of programs are available for manipulating WAV files as their digital representation is not complex. Recording sounds can be done at very high sample rates to generate very high-quality sounds, but the sample rate can be decided by the user doing the recording. However, WAV file sizes are generally too large to be practical for distribution online (due to bandwidth limitations) or mass storage on hand-held devices (where space is limited).

MP3

MP3 = MPEG Audio Layer 3*

The MP3 file format uses lossy compression techniques. Sounds that the human ear does not generally notice get discarded before other compression techniques get applied to eventually get the file size down to typically between one tenth and one eleventh of the original uncompressed file size. WAV files can be easily saved as files in MP3 format, enabling more sound data to fit into storage devices without a very noticeable deterioration in the sound quality. Music industry professionals tend not to use MP3 due to that deterioration being unacceptable.

WMA

WMA = Windows Media Audio

This is an audio file format developed by Microsoft to compete with MP3 for popularity. Like MP3, it uses lossy compression and the amount of compression can be varied.

---

* *You may also say "MPEG-1 Audio Layer 3", or "MPEG-2 Audio Layer 3"*

MIDI

MIDI = Musical Instrument Digital Interface

Music held in the MIDI format is captured via a serial interface connected to an electronic musical instrument or other sound system. The connections linking the two are called MIDI channels and data can be both sent to and received from the serial interface along MIDI channels. MIDI technology includes standardised MIDI codes that represent the instrument used, the tempo and the pitch, duration and volume of musical notes. Thanks to the MIDI standard, all MIDI technology is interoperable. MIDI is a format for the interchange of sound data and is effectively a series of commands instructing the computer how to play the sound back, but when sound is played back it is actual sampled recordings of sound clips that get played through the speaker. MIDI is thus not a format for the storage of sound data, merely its transmission.

Video File Types

MPEG

MPEG = Moving Picture Experts Group

MPEG = Motion Picture Experts Group

MPEG is a widely-used video file format (or, more accurately, a suite of formats). It uses delta compression which involves transmitting only the information about pixels that change between selected frames of a video. This significantly reduces the amount of data transmitted overall. To sustain the overall video quality, a whole frame will also be transmitted but at less frequent intervals. MPEG can be used on optical disks for various video resolutions and playback window sizes.

MP4

MP4 = MPEG-4 Part 14

The MP4 format was designed for digital streaming over the internet. It stores and plays digital audio with video. Since it forms part of the MPEG standards, it encompasses MP3 technology in its lower layers, therefore MP3 audio can be played on an MP4 player. It supports subtitling and online streaming, and the file can include still images. Lossy compression is used to reduce large video file sizes.

AVI

AVI = Audio Video Interleaved

AVI supports multiple audio and multiple video streams playing back from 1 file. No further compression is applied to existing clips saved as AVI format, which sustains their original quality and eliminates the need for codecs, but leads to higher file sizes until compression is (optionally) applied. AVI files do not store the ratio of width-to-height for playback.

Questions

Sample Question and Answer: GCSE ICT 2008 Paper 2 Q4(d)

Give one way a PDF document is different from a text-based file.

Allow any **one** of the following:
- PDF cannot be edited on screen [1]
- It is a smaller file [1]
- You need different software to open it [1]
- Can be opened on a wide variety of platforms [1]

[Source: CCEA GCSE ICT 2008 Paper 2 Question 4(d)]

Q1   Explain what is meant by data portability with reference to the CSV format.

Q2   (a)   State 1 difference between a PICT file and a GIF file.

      (b)   State 4 differences between a JPEG file and a GIF file.

Q3   Expand the following acronyms: PNG, AVI, PDF, MIDI, RTF, JPEG, MPEG, WMA

Q4   Here are the contents of the RTF document shown in the notes.
Identify ten RTF features used in the creation of the document.

Q5   Explain 3 reasons why PDFs are good for disseminating information globally.

Q6   How does an MPEG keep its file size low?

Q7   Unscramble the words to reveal 4 things supported by the MP4 file type. Clue: they all begin with S!

(a)   BTLINGUIST
(b)   LEGALSISTIM
(c)   DONUS
(d)   EMIGRANTS

Q8   Outline the key features of MIDI technology.

Q9   Outline two limitations of high-quality WAV recordings.

Q10   Explain why TXT files are highly portable.

Q11   Explain why images may need to be compressed before being uploaded for inclusion on a website.

Q12   Maria takes several photographs whilst on holiday.

(a)   Explain the issues she may encounter with storing a large number of images.

(b)   Discuss three ways in which she may resolve the issues, evaluating each solution.

Bibliography

BCS Glossary of Computing and ICT, 13th ed., BCS Academy Glossary Working Party

https://www.prepressure.com/library/file-formats/pict [Accessed 8/1/17]

http://whatis.techtarget.com/fileformat/TXT-ASCII-text-formatted-data [Accessed 8/1/17]

http://www.thoughtspot.com/blog/6-rules-creating-valid-csv-files [Accessed 8/1/17]

http://searchexchange.techtarget.com/definition/Rich-Text-Format [Accessed 8/1/17]

http://www.simpopdf.com/resource/pdf-features.html [Accessed 8/1/17]

http://www.dawsons.co.uk/blog/how-do-mp3-and-wav-files-differ [Accessed 8/1/17]

http://www.huffingtonpost.ca/penny-will/the-wonder-of-midi-files_b_1823475.html [Accessed 9/1/17]

http://mpeg.chiariglione.org/ [Accessed 9/1/17]

CCEA ICT for GCSE Student's Book (2011), S. Matthewson, G. Lynch & M. Debbadi

http://www.makeuseof.com/tag/technology-explained-what-is-the-difference-between-mp3-mp4/ [Accessed 9/1/17]

https://www.winxdvd.com/resource/avi.htm [Accessed 9/1/17]