

## Chapter 26

# Introduction to Multimedia

### Objectives

- What is Multimedia?
- The Sound formats
- The Video formats
- Playing Sound on the Web
- Playing Video on the Web
- Speech synthesis & recognition

**Multimedia is pictures, sounds, music, animations and videos.**

**Modern web browsers have support for many multimedia formats.**

### What is Multimedia?

Multimedia is everything you can hear or see: texts, books, pictures, music, sounds, CDs, videos, DVDs, Records, Films, and more.

Multimedia comes in many different formats. On the Internet you will find many of these elements embedded in web pages, and today's web browsers have support for a number of multimedia formats.

In this tutorial you will learn about different multimedia formats and how to use them in your web pages.

In today's world we are often faced with vast streams of information. We process this information both consciously and unconsciously. Lets take a tour of a department store, for example. A guide map at the entrance of this store tells you about the location of different departments and what is available at each counter. Each section in the store has an attractive display of the various items. Soft background music offers the customers a subtle persuasion to buy. A helpful salesperson gives detailed information about each product.

After reading the above description it might have already occurred to you that the department store has very thoughtfully used different media elements to convey important information about the wide cross-section of items available in the store. The magic thus created by the right mixing of different media elements would have been lost if either one of these elements had been missing.

The most pertinent question that will be tackled in this chapter is What is Multimedia? Well here's a simple explanation:

"Multimedia is the incorporation of sound, animation, still images, hypertext and video integrated under the control of a computer program."

Multimedia may use some or all of these aspects of communication. An important point to be noted is that multimedia is not a product, it is a technology (or more accurately a combination of technologies). Technology in simple terms is anything that makes something else more efficient. Multimedia aims at making computers easier to use.

## **Exploring Multimedia**

With multimedia you can explore the sights and sounds of the Jurassic era, study the music and ideas of your favorite composer; or learn a new language by listening to its pronunciation and practicing conversation. Multimedia opens up a world of creativity, letting you add video and sound to your documents - or even produce your own electronic book, complete with sound and animation!

Some multimedia programs are linear, i.e., a sort of slide show on a computer screen, in which user involvement is limited to advancing to the next information screen or going back to view the previous one.

Interactive Multimedia is another buzzword going through the industry - this is the second kind of multimedia program. This is the key element of multimedia concept, where in the user is allowed to take control of program execution.

The books we read are not very interactive, and neither is the television we watch. With the inclusion of a keyboard and a mouse, a computer can easily accept user input. This is what makes PC's, the premier interactive multimedia machines of today.

In a totally interactive multimedia concept, the user can influence the flow of information. Clicking on a predefined area known as a hotspot will display another file in the program; the linked file can be a image, sound file, a digital video clip, or a window filled with new information.

Now, let's observe the following pictures carefully, note the continuity among them, doesn't the obvious link in them strike you ? The pictures depict what a typical multimedia application would be like. The initial screen gives you a broad information on what is to follow, and by use of the different media components the entire story unfolds its self to you.

## **What does Multimedia do?**

Multimedia can certainly be very impressive - you can run digital video in a window on your desktop and have music coming out of your PC speakers. It is but natural to think that if multimedia combines sound, video, graphics and text; how is it different from a TV?

Every night, the news you can see is a combination of all of these elements. The only missing link is interactivity, you have no control on your TV; save change channels!. A multimedia version of a newscast could contain the entire newscast, but you will be the one who decides what information should be narrated by the newscaster, and in what sequence. The choice totally rests with you. Just click a button touches a screen, or press a key, and you will be the one who calls the shots within a multimedia application.

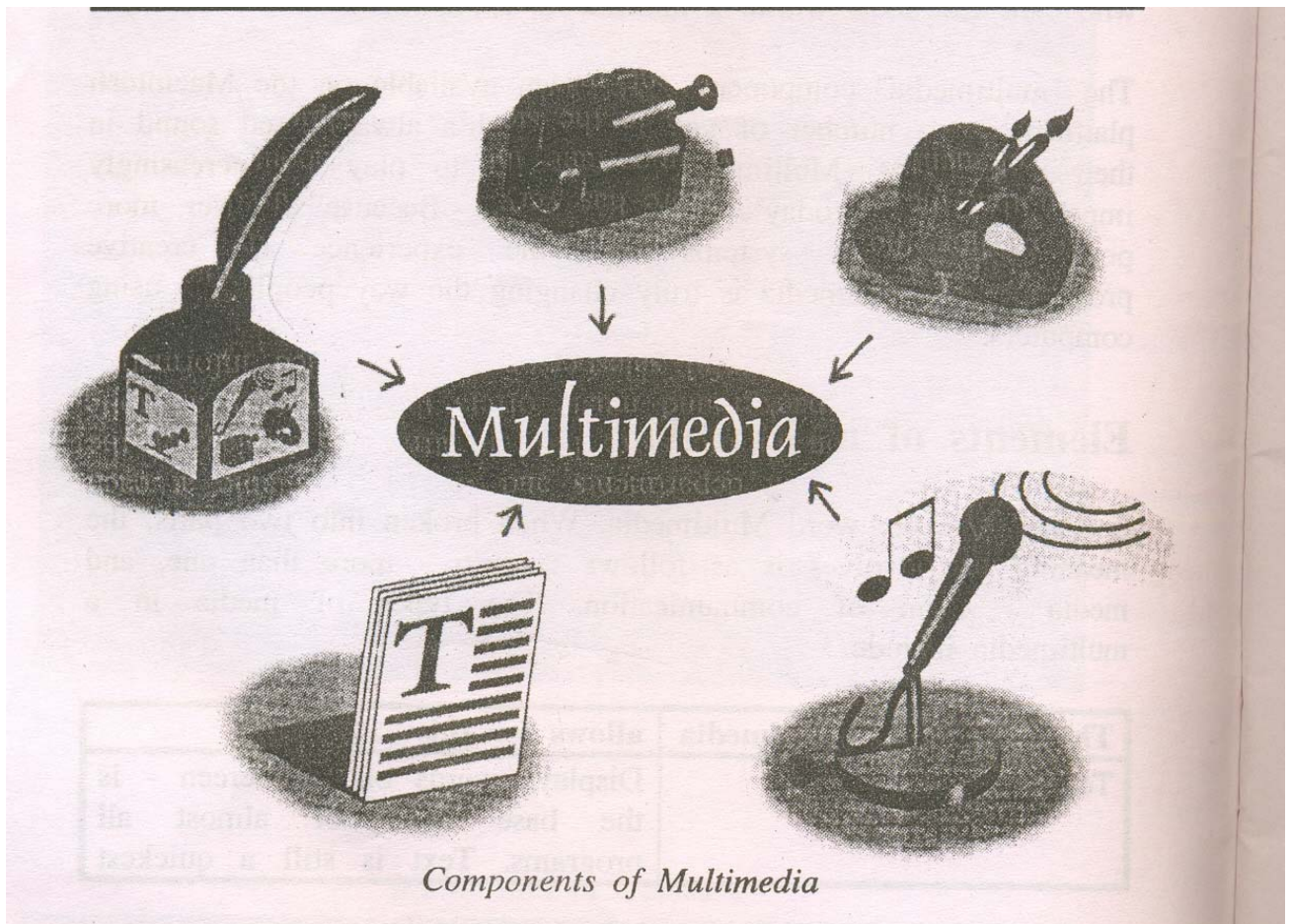
The "multimedia" components have been available on the Macintosh platform for a number of years. Macintosh's always used sound in their applications. Multimedia has started to play an increasingly important role in today's computer world. Because of ever more

- powerful computer systems and the experience of creative
- programmers, multimedia is truly changing the way people are using
- computers.

## Elements of multimedia

Lets examine the word Multimedia. When broken into two parts, the meaning that evolves is as follows: multi - more than one, and media - form of communication. The types of media in a multimedia include:

Text	Displays word on the screen – is the base layer of almost all programs. Text is still the quickest programs will always use it.
Pictures	Multimedia computers can display photograph quality pictures on the monitor.
Movies	With a multimedia program, your computer can turn into a TV set.
Animation	Sometimes animation can express a point better than movies.
Sound	A biting dinosaur isn't much impressive unless you can hear the bone crunches well.



## **Multimedia requirements**

The basic idea of "multimedia" is to manage and co-ordinate various devices of communications and entertainment electronics with the PC as a central controller.

Using a few additional pieces of hardware most standard PCs can control multimedia applications. The appropriate multimedia software is required to combine and integrate the individual component video, music, animation, graphics and text according to user requirements. An IBM-compatible PC has several features specific to multimedia display. A multimedia system is a combination of mainly three components:

- hardware devices that serve as an interface to digital computers and analog devices.
- computer hardware that enables the processing and management of multimedia information.
- software to process and integrate the information. Example, drivers, editors, processing software and authoring software.

All multimedia systems may not require all these hardware and software. By itself multimedia is nothing new. Recorded sound, movies and pictures have been around for years. The new part is the way computers can intertwine these things.

## **Browser Support**

The first Internet browsers had support for text only, and even the text support was limited to a single font in a single color, and little or nothing else.

Then came web browsers with support for colors, fonts and text styles, and the support for pictures was added.

The support for sounds, animations and videos is handled in different ways by different browsers. Some elements can be handled inline, some requires a plug-in and some requires an ActiveX control.

## **Multimedia Formats**

Multimedia elements (like sounds or videos) are stored in media files. The most common way to discover the media type is to look at the file extension.

When a browser sees the file extensions .htm or .html, it will assume that the file is an HTML page. The .xml extension indicates an XML file, and the .css extension indicates a style sheet.

Picture formats are recognized by extensions like .gif and .jpg.

Multimedia elements also have their own file formats with different extensions.

## **Multimedia Sound Formats**

**Sound can be stored in many different formats.**

### **The MIDI Format**

The MIDI (Musical Instrument Digital Interface) is a format for sending music information between electronic music devices like synthesizers and PC sound cards.

The MIDI format was developed in 1982 - long before the Internet -, but since it was developed by the music industry, it is very flexible and can be used for everything from very simple to real professional music making.

MIDI files do not contain sampled sound, but a set of digital musical instructions (musical notes) that can be interpreted by your PC's sound card.

The downside of MIDI is that it cannot record sounds (only notes). Or, to put it another way: It cannot store songs, only tunes.

The upside of the MIDI format is that since it contains only instructions (notes), MIDI files can be extremely small. The MIDI format is supported by many different software systems over a large range of platforms. MIDI files are supported by all the most popular Internet browsers.

Sounds stored in the MIDI format have the extension .mid or .midi.

### **The RealAudio Format**

The RealAudio format was developed for the Internet by Real Media. The format also supports video.

The format allows streaming of audio (on-line music, Internet radio) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.

Sounds stored in the RealAudio format have the extension .rm or .ram.

### **The AU Format**

The AU format is supported by many different software systems over a large range of platforms.

Sounds stored in the AU format have the extension .au.

### **The AIFF Format**

The AIFF (Audio Interchange File Format) was developed by Apple. AIFF files are not cross-platform and the format is not supported by all web browsers.

Sounds stored in the AIFF format have the extension .aif or .aiff.

### **The SND Format**

The SND (Sound) was developed by Apple.

SND files are not cross-platform and the format is not supported by all web browsers.

Sounds stored in the SND format have the extension .snd.

### **The WAVE Format**

The WAVE (waveform) format is developed by IBM and Microsoft.

It is supported by all computers running Windows, and by all the most popular web browsers.

Sounds stored in the WAVE format have the extension .wav.

### **The MP3 Format (MPEG)**

MP3 files are actually MPEG files. But the MPEG format was originally developed for video by the Moving Pictures Experts Group. We can say that MP3 files are the sound part of the MPEG video format.

MP3 is one of the most popular sound formats for music recording. The MP3 encoding system combines good compression (small files) with high quality. Expect all your future software systems to support it.

Sounds stored in the MP3 format have the extension .mp3, or .mpga (for MPG Audio).

### **What Format To Use?**

The WAVE format is one of the most popular sound format on the Internet, and it is supported by all popular browsers. If you want recorded sound (music or speech) to be available to all your visitors, you should use the WAVE format.

The MP3 format is the new and upcoming format for recorded music. If your website is about recorded music, the MP3 format is the choice of the future.

### **Multimedia Video Formats**

**Video can be stored in many different formats.**

#### **The AVI Format**

The AVI (Audio Video Interleave) format was developed by Microsoft.

The AVI format is supported by all computers running Windows, and by all the most popular web browsers. It is a very common format on the Internet, but not always possible to play on non-Windows computers.

Videos stored in the AVI format have the extension .avi.

#### **The MPEG Format**

The MPEG (Moving Pictures Expert Group) format is the most popular format on the Internet. It is cross-platform, and supported by all the most popular web browsers.

Videos stored in the MPEG format have the extension .mpg or .mpeg.

#### **The QuickTime Format**

The QuickTime format is developed by Apple. QuickTime is a common format on the Internet, but QuickTime movies cannot be played on a Windows computer without an extra (free) component installed.

Videos stored in the QuickTime format have the extension .mov.

## **The RealVideo Format**

The RealVideo format was developed for the Internet by Real Media. The format allows streaming of video (on-line video, Internet TV) with low bandwidths. Because of the low bandwidth priority, quality is often reduced.

Videos stored in the RealVideo format have the extension .rm or .ram.

## **The Shockwave (Flash) Format**

The Shockwave format was developed by Macromedia. The Shockwave format requires an extra component to play. This component comes preinstalled with the latest versions of Netscape and Internet Explorer.

Videos stored in the Shockwave format have the extension .swf.

## **Playing Sounds On The Web**

**Sounds can be played "inline" or by a "helper", depending on the HTML element you use.**

### **Inline Sound**

When sound is included in a web page, or as part of a web page, it is called inline sound. Inline sound can be added to a web page by using the <bgsound> element or the <img> element.

If you plan to use inline sounds in your web applications, be aware that many people find inline sound annoying. Also note that some users might have turned off the inline sound option in their browser.

Our best advice is to include inline sound only in web pages where the user expects to hear the sound.

### **Using A Helper (Plug-In)**

A helper application is a program that can be launched by the browser to "help" playing sound. Helper applications are also called Plug-Ins.

Helper applications can be launched using the <embed> element, the <applet> element, or the <object> element.

One great advantage of using a helper application is that you can let some (or all) of the player settings be controlled by the user.

Most helper applications allow manually (or programmed) control over the volume settings and play functions like rewind, pause, stop and play.

### **Using The <bgsound> Element**

Internet Explorer supports an element called <bgsound>.

The purpose of this element is to provide a background sound for a web page:

```
<bgsound src="beatles.mid" />
```

The code fragment above displays a MIDI file as background music for a web page. A list of attributes for the <bgsound> element can be found in a later chapter of this tutorial.

The <bgsound> element is not a standard HTML or XHTML element. It is supported by Internet Explorer only.

#### Using The <img> Element

Internet Explorer supports the dynsrc attribute in the <img> element.

The purpose of this element is to embed multimedia elements in web page:

```

```

The code fragment above displays a WAVE file embedded in a web page.

The dynsrc attribute is not a standard HTML or XHTML attribute. It is supported by Internet Explorer only.

#### Using The <embed> Element

Internet Explorer and Netscape both support an element called <embed>.

The purpose of this element is to embed multimedia elements in web page:

```
<embed src="beatles.mid" />
```

The code fragment above displays a MIDI file embedded in a web page. A list of attributes for the <embed> element can be found in a later chapter of this tutorial.

The <embed> element is supported by both Internet Explorer and Netscape, but it is not a standard HTML or XHTML element. The World Wide Web Consortium (W3C) recommends using the <object> element instead.

#### Using The <object> Element

Internet Explorer and Netscape both support an HTML element called <object>.

The purpose of this element is to embed multimedia elements in web page:

```
<object  
classid="clsid:22D6F312-B0F6-11D0-94AB-0080C74C7E95">  
<param name="FileName" value="liar.wav" />  
</object>
```

The code fragment above displays a WAVE file embedded in a web page. A list of attributes for the <object> element can be found in a later chapter of this tutorial.

#### Using A Hyperlink

If a web page includes a hyperlink to a media file, most browsers will use a "helper application" to play the file:

```
<a href="beatles.mid">  
Click here to play the Beatles  
</a>
```



The code fragment above displays a link to a MIDI file. If the user clicks on the link, the browser will launch a helper application like Windows Media Player to play the MIDI file.

## Playing Videos On The Web

**Videos can be played "inline" or by a "helper", depending on the HTML element you use.**

### Inline Videos

When a video is included in a web page it is called inline video.

Inline video can be added to a web page by using the <img> element.

If you plan to use inline videos in your web applications, be aware that many people find inline videos annoying. Also note that some users might have turned off the inline video option in their browser.

Our best advice is to include inline videos only in web pages where the user expects to see a video. An example of this is a page which opens after the user has clicked on a link to see the video.

### Using A Helper (Plug-In)

A helper application is a program that can be launched by the browser to "help" playing a video. Helper applications are also called Plug-Ins.

Helper applications can be launched using the <embed> element, the <applet> element, or the <object> element.

One great advantage of using a helper application is that you can let some (or all) of the player settings be controlled by the user.

Most helper applications allow manual (or programmed) control over the volume settings and play functions like rewind, pause, stop and play.

### Using The <img> Element

Internet Explorer supports the dynsrc attribute in the <img> element.

The purpose of this element is to embed multimedia elements in web page:

```

```

The code fraction above displays an AVI file embedded in a web page.

The dynsrc attribute is not a standard HTML or XHTML attribute. It is supported by Internet Explorer only.

### Using The <embed> Element

Internet Explorer and Netscape both support an element called <embed>.

The purpose of this element is to embed multimedia elements in web page:

```
<embed src="video.avi" />
```

The code fraction above displays an AVI file embedded in a web page.

A list of attributes for the <embed> element can be found in a later chapter of this tutorial.

The <embed> element is supported by both Internet Explorer and Netscape, but it is not a standard HTML or XHTML element. The World Wide Web Consortium (W3C) recommend using the <object> element instead.

#### Using The <object> Element

Internet Explorer and Netscape both support an HTML element called <object>.

The purpose of this element is to embed multimedia elements in web page:

```
<object data="video.avi" type="video/avi" />
```

The code fraction above displays an AVI file embedded in a web page.

A list of attributes for the <object> element can be found in a later chapter of this tutorial.

#### Using A Hyperlink

If a web page includes a hyperlink to a media file, most browsers will use a "helper application" to play the file:

```
<a href="video.avi">  
Click here to play a video file  
</a>
```

The code fraction above displays a link to an AVI file. If the user clicks on the link, the browser will launch a helper application like Windows Media Player to play the AVI file.

### **Speech recognition and synthesis**

Much of speech recognition and synthesis is SP, and these remain active research areas. Existing recognition algorithms remain constrained by artificial grammars or limited vocabularies and usually require long training for individual speakers.

### **Introduction**

This article will look at support for speech in Microsoft Windows and see what's involved in incorporating aspects of speech technology in Windows applications. In particular we examine the Microsoft Speech API (SAPI) to see what it offers developers in terms of letting applications speak to users and also understand what users say. Since there is a lot of information (text and code) here, the article has been split over a number of pages. This page is an introduction to the subject whilst the other two pages (which have enough information to make them individual articles) look in detail at using SAPI 4 and SAPI 5.1 within Delphi applications.

Once upon a time the task of making your application speak or understand its user's commands was science fiction or at the very least involved lots of computing power. We will see that today's computers and speech technology enable any application to use speech technology with relative ease and good performance.

Microsoft have been researching and implementing speech technology for some years and they have an area of their Web site dedicated to the matter at <http://www.microsoft.com/speech>.

## **Speech Technology**

The speech capabilities that can be added to an application are text-to-speech synthesis (TTS) and speech recognition (SR).

### **Text-To-Speech Synthesis (TTS)**

This involves turning a string into spoken language that is played through the computer speakers. The complexities of turning words into phonemes, adding appropriate emphasis and translating the result into digital audio are beyond the scope of this paper and are catered for by a TTS engine installed on your machine.

The end result is that the computer talks to the user to save the user having to read some text on the screen.

### **Speech Recognition (SR)**

This involves the computer taking the users speech and interpreting what has been said. This allows the user to control the computer (or certain aspects of it) by voice, rather than having to use the mouse and keyboard, or alternatively just dictating the contents of a document.

The complex nature of translating the raw audio into phonemes involves a lot of signal processing and is not focused on here. These details are taken care of by an SR engine that will be installed on your machine. SR engines are often called *recognizers* and these days typically implement continuous speech recognition (older recognizers implemented isolated or discrete speech recognition, where pauses were required between words).

Speech recognition usually means one of two things. The application can understand and follow simple commands that it has been educated about in advance. This is known as *command and control* (sometimes seen abbreviated as CnC, or simply SR).

Alternatively an application can support *dictation* (sometimes abbreviated to DSR). Dictation is more complex as the engine has to try and identify arbitrary spoken words, and will need to decide which spelling of similarly sounding words is required. It develops context information based on the preceding and following words to try and help decide. Because this context analysis is not required with Command and Control recognition, CnC is sometimes referred to as context-free recognition.

## **Speaker Profiles**

Dictation speech recognition is speaker-dependant, meaning that because of different people's enunciation, accent, pitch and so on, recognizers require a speaker profile to be set up for decent results. This profile results from training sessions that educate the recognizer about the nuances of the speaker's voice.

On the other hand, command and control speech recognition is usually not speaker-independent.

## **The Previous State Of Affairs**

As the incorporation of speech technology became more realistic more vendors released TTS and SR engines. Unfortunately each engine had its own API and so interchanging them was not possible. Programming for multiple engines meant a lot of recording and the whole situation was very similar to the database API programming problem before the advent of the BDE and ADO.

## **The Current State Of Affairs**

In late 1995 the Microsoft Speech API (SAPI) was introduced as part of the Windows Open Services Architecture (WOSA) services. This was intended to simplify matters and has done a good job of doing so, at least relatively speaking. Depending on what you wish to do it can involve some tricky coding, but that's the same story with the basic Windows API.

SAPI is currently (at the time of writing) at version 5.1 and now professes a single API made from a set of interfaces that you can program it with to get TTS and/or SR in your application. However up until version 4 an alternative API was in use. In fact there were two APIs defined, but neither of these is now documented and Microsoft recommends the new API.

This means that TTS and SR engines will have to be labeled as either SAPI 4 compliant (if they use the old interfaces) or SAPI 5 compliant (if they use the new interfaces). However, given the widespread use of the older interfaces you shouldn't expect Microsoft to stop them being available any time soon.

The Microsoft Speech SDK can be obtained via Microsoft's Web site and when installed provides documentation on the APIs. Because of the complete differences between SAPI 4 and SAPI 5.1 you can install both SDKs on a single machine and take advantage of any of the available APIs.

SAPI applications programmers call the interfaces defined in the API and SAPI-compliant TTS and SR engines implement those interfaces. SAPI supports text to speech (TTS), speech recognition (SR), dictation speech recognition (DSR) and also telephony (TEL). We will explore TTS and also see what's involved with both types of SR in this paper.

## **Summary**

Multimedia is pictures, sounds, music, animations and videos. Modern web browsers have support for many multimedia formats. The basic idea of "multimedia" is to manage and co-ordinate various devices of communications and entertainment electronics with the PC as a central controller. The MIDI (Musical Instrument Digital Interface) is a format for sending music information between electronic music devices like synthesizers and PC sound cards. The AVI format is supported by all computers running Windows, and by all the most popular web browsers. In particular we examined the Microsoft Speech API (SAPI) to see what it offers developers in terms of letting applications speak to users and also understand what users say.

## **Questionnaire**

1. Explain in brief about Multimedia.
2. Give the sound & video format.
3. Write about speech recognition & synthesis.
4. How to play sound on web pages?
5. How to play videos on web pages?
6. What are the requirements of multimedia?