

Digital Dictation Buzzwords Guide

The following guide has been compiled on the basis of a general search of digital dictation vendor's web sites with a view to demystifying and explaining some of the buzzwords used in relation to digital dictation. The definitions of the words included here have in general been extracted from various Internet sources and by no way constitute a definitive explanation, instead this is more a guide to the technical terms commonly used so that they can be considered when evaluating products. We hope it is of use.

A	
ACELP	'Algebraic Code Excited Linear Prediction'. This acronym describes the algorithm used in this lossy audio compression process (see 'lossy compression'). It uses a codebook developed from mathematical modelling of the vocal tract to encode and decode compressed audio.
ADPCM	'Adaptive differential pulse-code modulation' is a CODEC for converting sound information to binary information (a string of 0's and 1's) by taking frequent samples of the sound and expressing the value of the sampled sound modulation in binary terms.
Analogue	An analogue signal such as audio can be represented as a series of sine waves. The term originated because the modulation of the carrier wave is analogous to the fluctuations of the human voice or other sound that is being transmitted. Traditional dictation tape machines are considered to be analogue devices.
API	'Application Programming Interface'. An application program interface is the specific method prescribed by a computer operating system or by an application program by which a programmer writing an application program can make requests of the operating system or another application.
Application Server	An application server is a server program in a computer in a distributed network that provides the business logic for an application program. The application server is frequently viewed as part of a three-tier application (see 'Three-Tier').
Architecture	In information technology, especially computers and more recently networks, architecture is a term applied to both the process and the outcome of thinking out and specifying the overall structure, logical components, and the logical interrelationships of a computer, its operating system, a network, or other components.
Archive/Archiving	An archive is a collection of computer files that have been packaged together for backup, to transport to some other location, for saving away from the computer so that more hard disk storage can be made available, or for some other purpose. Archiving is the process whereby an application stores files in an archive.
Audio	This is the part of digital dictation where methods of sound recording and playback are employed to capture and playback the author's voice in a digital format of 1 and 0s.
Auto Saving	A process whereby a dictation that is being worked on can be recovered in case of a PC crash or system failure. It can be implemented in different ways where it is done transparently or the user is interrupted whilst this happens. Again systems that have implemented this vary in the amount of recorded sound they recover from 100% to everything but the last 2 seconds.

B	
Bandwidth	<p>Bandwidth (the width of a band of electromagnetic frequencies) is used to describe how fast data flows on a given transmission path.</p> <p>Generally speaking, bandwidth is directly proportional to the amount of data transmitted or received per unit time. In a qualitative sense, bandwidth is proportional to the complexity of the data for a given level of system performance. For example, it takes more bandwidth to download a photograph in one second than it takes to download a page of text in one second. Large sound files, computer programs, and animated videos require still more bandwidth for acceptable system performance.</p>
Batch Recognition	<p>The technique of sending sound files to a remote server where they are processed by a speech recognition engine in to text before the text and the original soundfile are returned for correction of recognition errors and completion of the document. This is typically configured where an author records a sound file, sends it for recognition and then a secretary corrects the resulting text.</p>
Bits per Sample	<p>The instantaneous amplitude of an audio signal at each sampling (see 'Sample Rate') is rounded off to the nearest of several specific, predetermined levels. This process is called quantisation. The number of levels is always a power of 2 -- for example, 8, 16, 32, or 64. These numbers can be represented by three, four, five, or six binary digits (bits) respectively. The digital representation of an audio waveform is thus a series of binary numbers, each represented by some power of 2 bits. The lower the bits per sample the smaller the resulting soundfile will be however the quality of sound on play back will also be lower.</p>
Bookmarks	<p>Markers that can be placed in a dictation to remember a position so that it can be quickly returned to at a later date. Commonly known in traditional analogue tape machines as 'index' marks.</p>
Bytes per Second or Kilobytes per Second (Kb/sec)	<p>This is the average data transfer rate for a CODEC and gives the best indication of what bandwidth will be used by a particular scheme when streaming or what the resulting soundfile size will be. This is a function of the sampling rate and the bits per sample for the CODEC.</p>

C	
CD Quality	<p>This is a subjective term for the recording of sound in the PCM 44.1Khz 16bit stereo format (See 'PCM'). It typically results in a 10min dictation file being approximately 100MB in size.</p>
CELP	<p>'Code Excited Linear Prediction'. This acronym describes the algorithm used in this lossy audio compression process (see 'lossy compression'). It uses a codebook developed from mathematical modelling of the vocal tract to encode and decode compressed audio.</p>
Client	<p>A client is the requesting program or user in a client/server relationship. (See 'Client/Server')</p>
Client/Server	<p>Client/server describes the relationship between two computer programs in which one program, the client, makes a service request from another program, the server, which fulfils the request. Although the client/server idea can be used by programs within a single computer, it is a more important idea in a network. In a network, the client/server model provides a convenient way to interconnect programs that are distributed efficiently across different locations.</p>
Client Side Workflow	<p>This is where the workflow is implemented within each desktop client application. Whilst the traditional way of producing simple workflow solutions it has the disadvantage that if the workflow is updated every desktop needs to be visited to apply this update. (See 'Server Side Workflow').</p>
CODEC	<p>The term codec is also an acronym that stands for 'compression/decompression.' A codec is an algorithm, or specialized computer program, that reduces the number of bytes consumed by large files and programs. Audio compression is typically a lossy compression process (see 'lossy compression').</p>

Compression	Compression is the reduction in size of data in order to save space or transmission time. In terms of dictation this would be a recorded audio data. Compression is performed by a program that uses a formula or algorithm to determine how to compress or decompress data.
Context	Another term commonly used for a speech recognition vocabulary. See 'Vocabulary'.
CTI	'Computer Telephony Interface'. This is the use of dedicated PC hardware and software to connect a PC to a telephone line so that the PC can respond and handle incoming and outgoing calls. Often used in digital dictation to provide a telephone dictation interface.

D

Database	A database is a collection of data that is organized so that its contents can easily be accessed, managed, and updated. The most prevalent type of database is the relational database, a tabular database in which data is defined so that it can be reorganized and accessed in a number of different ways. Most digital dictation workflow systems use a relational database to track and route dictations and the associated information such date submitted, length, priority between users. The database also provides audit information on the dictations showing where they have been and how long they took to get there.
Database Server	This is a server (see 'Server') that has a database (see 'Database') engine running on it and provides database services to clients (see 'Clients').
Dial-Up	This is method of connecting two computers together so that they can share files or a remote computer to network so that it can access data and services provided on that network. Typically this is done using a remote access server (see 'RAS') and a direct dial telephone line.
Digital Pocket Memo	A brand name typically associated with the Philips range of dictating equipment, in particular the Philips DPM9300 Digital Pocket Memo
Distributed Server	This is a system architecture whereby all server services (see 'Server Services') can be distributed across multiple servers (see 'Servers') eliminating a risk of single point of failure (see 'Single Point of Failure'). Where a system employs a Three-tier on n-Tier (see 'Three-Tier' and 'n-Tier') architecture it is not considered to be truly distributed if it only allows some server services to be run across multiple servers as the single point of failure risk still exists.
DSP Truespeech	DSP Group's TrueSpeech is a family of high quality, low bit rate, speech compression CODECS (see 'CODEC') which compress speech down to as little as 1/40th its original size. Several different versions of TrueSpeech at different compression rates are available for licensing, from 8.5 Kbps through 3.9 Kbps.
DSS file	'Digital Speech Standard' file. A proprietary speech compress CODEC (see 'CODEC') and audio data file format developed by the International Voice Association (IVA). Usually requires proprietary software to playback these files. Under its licensing terms it can only be used for recording when used on portable recording devices.
DTMF	DTMF (dual tone multi frequency) is the tone you generate when you press an ordinary telephone's touch keys. With DTMF, each key you press on your phone generates two tones of specific frequencies. So that a voice can't imitate the tones, one tone is generated from a high-frequency group of tones and the other from a low frequency group. The tones, generated as a user presses a key in response to a prompt, are typically used to control the flow of telephone voice applications such as telephone dictation or voice mail

E

F	
Fat Client	A 'Fat Client' is a fully configured PC with multiple applications installed. It is the opposite to a 'Thin Client' (see 'Thin Client'). This is the typical configuration of most office PCs
File Server	A file server is a server (see 'Server') that provides file read/write access and storage for clients (see' Clients') on a network.
G	
GSM6.10	'Global System for Mobile communications'. This is a lossy (see 'lossy compression') voice codec most commonly used in digital cellular communications. It can deliver a quality suitable for transcription with very small data transfer rates (see 'Bytes per second').
H	
I	
Internet Client	A true internet client is typically considered to be a client application that runs on a web application server remote of the PC where the users is working. Vendors do however often tenuously refer to versions of their product as an 'Internet Client' when the application runs locally and simply uses the internet as an extension of the network on which their target server lives.
Inserting	This is the novel feature available in most digital dictation applications that allows the insertion of sound at any point in a dictation. Similar to inserting text in a word processor, as the sound is recorded the existing recorded sound is moved apart to allow room for the new recording.
J	
K	
L	
LAN	'Local Area Network'. A local area network is a group of computers and associated devices that share a common communications line attaching to one or more servers within a single geographical location
Load Balancing	Load balancing is dividing the amount of work that a computer has to do between two or more computers so that more work gets done in the same amount of time and, in general, all users get served faster. Load balancing can be implemented with hardware, software, or a combination of both. Typically, load balancing is the main reason for computer server clustering. Usually, if two servers are used to balance a work load, a third server is needed to determine which server to assign the work to.
Lossless Compression	In lossless compression (see 'Compression'), every single bit of data that was originally in the file remains after the file is uncompressed. All of the information is completely restored. This is generally the technique of choice for text or spreadsheet files, where losing words or financial data could pose a problem.
Lossy Compression	Lossy compression (see 'Compression') reduces a file by permanently eliminating certain information, especially redundant information. When the file is uncompressed, only a part of the original information is still there (although the user may not notice it). Lossy compression is generally used for video and sound, where a certain amount of information loss will not be detected by most users.

LPC 'Linear Predictive Coding' is a CODEC (see 'CODEC') algorithm that can be used for the compression of speech audio.

M

MSDE 'Microsoft Database Engine' is a data engine built and based on core Microsoft SQL Server technology. This royalty free product allows small practices to take advantage of client/server applications without needing to purchase a full version of MS SQL Server. Microsoft advises that this product contains a scalability (see 'Scalability') limiter and performance degrades when 5 or more concurrent SQL transactions are carried out. Most digital dictation vendors have found this equates to around 100 users using the system although this is dependant upon the way the vendor's application talks to the database. MSDE can be upgraded to full MS SQL Server at any time.

N

O

Offline Recognition See 'Batch Recognition'

Offline Working This is the process whereby a client (see 'Client') application allows normal working without needing to attach to a server.

Online Recognition This is typically used to describe the process of using computer with speech recognition to produce text on a screen as the user speaks. It is the most commonly seen form of speech recognition.

Online Working This is the process whereby a client (see 'Client') application requires or uses the services of a server during normal working.

Outsourced Typing This is the sending of transcription to an external agency for typing. With the proliferation of digital dictation as well as other electronic office automation these services are growing in number and the services they offer.

Overwriting This is done during dictation when recording over the top of a previously recorded section and is typically the way dictation tape recorders work. Using the analogy of a word processor this is like when you type and the text you produce writes over the existing text. Most solutions provide this feature as an option to inserting (see 'Inserting').

P

PCM 'Pulse Code Modulation'. Originally developed in the 1940's PCM (pulse code modulation) is a digital scheme for transmitting analogue data. The signals in PCM are binary; that is, there are only two possible states, represented by logic 1 (high) and logic 0 (low). PCM is typically the way PC sound cards digitise sound data before it is compressed using the various audio compression schemes and is sometimes referred to as RAW audio.

Q

Queue In computer technology, a queue is a sequence of work objects that are waiting to be processed. In digital dictation the dictations that have been submitted by the authors are often referred to as the 'job queue' in that the dictations are waiting to be processed by a secretary or secretaries.

R	
RAS (Remote Access Server)	A remote access server is the computer and associated software that is set up to handle users seeking access to network remotely. Sometimes called a communication server, a remote access server usually includes or is associated with a firewall server to ensure security and a router that can forward the remote access request to another part of the corporate network.
Redundancy	This term is generally used to indicate that a system has internal processes that allow elements of the system to fail without causing significant loss of service. . Due to the subjective nature in which this term is often used it is always recommended that a vendors claims are substantiated with technical detail. (See 'Resilience').
Remote Working	This is the method of using a client application (see 'Client') with a RAS dial-up (see 'RAS') or VPN (see 'VPN') connection or an internet client (see 'Internet Client') to work effectively away from the office. This could be either at a clients offices or at home.
Resilience	This term is generally used to indicate the system is considered resilient to the failure of certain elements. Due to the subjective nature in which this term is often used it is always recommended that a vendors claims are substantiated with technical detail. See 'Redundancy'.
S	
Sample Frequency	To obtain a digitised form from an audio waveform the analogue signal amplitude is sampled (measured) at regular time intervals. The sampling rate, or number of samples per second, is several times the maximum frequency of the analogue waveform in cycles per second or hertz. The lower the sampling rate the smaller the resulting soundfile will be however the quality of sound on play back will also be lower.
Scalability	<ol style="list-style-type: none"> (1) It is the ability of a computer application or product (hardware or software) to continue to function well over the long term as it (or its context) is changed in size or volume in order to meet a user need. (2) It is the ability to not only function well but also take full advantage of scaling. For example, an application program would be scalable if it could be moved from a smaller to a larger operating system and take full advantage of the larger operating system in terms of performance (user response time and so forth) and the larger number of users that could be handled. <p>Whilst some systems claim scalability by reference it is not believed this implies proof of scalability as the system may not be taking full advantage of the overall infrastructure to deliver better performance, resilience or redundancy (see 'Resilience' and 'Redundancy'). Detailed investigation of the technology used is always recommended.</p>
SDK	Software Developers Kit. A software developer's kit (SDK) is a set of programs or components built around a product and used by a computer programmer to write application programs or integrations between applications. Traditionally seen as a set of Application Program Interfaces (API) this can also be in the form of ActiveX COM components, which allow faster and easier application development.
Server	<ol style="list-style-type: none"> 1) In general, a server is a computer program that provides services to other computer programs in the same or other computers. 2) The computer that a server program runs in is also frequently referred to as a server (though it may contain a number of server and client programs). 3) In the client/server (see 'Client/Server') model, a server is a program that awaits and fulfils requests from client programs (see 'Client') in the same or other computers. A given application in a computer may function as a client with requests for services from other programs and also as a server of requests from other programs.
Server Services	These are services provided by a server (see 'Server'). They can be in the form of file access and storage, database access or business logic programs.

Single Point of Failure	This is a term used to indicate the possibility that the failure of a single element in a system could cause the whole system not to operate. For example this could be a system that uses only one database instead of a number (see 'Distributed Server' and 'Single Server'). If this database failed then the whole system would cease to function.
Single Server	This is a system architecture whereby there is only one server for a primary service such as database. It is efficient to administrate but does introduce an inherent risk of a single point of failure.
Server Side Workflow	This is where the workflow implemented in a solution is generated using procedures that are stored on the server rather than being embedded in the clients (See 'Client/Server'). This has benefit that if the workflow is updated or improved only the server need be updated rather than every single desktop. This significantly reduces the maintenance requirement of a product. (See 'Client Side Workflow')
SoundBlaster	This is the brand name for soundcards manufactured by Creative Labs. Whilst few standards exists in soundcard manufacture 'SoundBlaster 16 Compatible' is a generic term often used to imply a sort of standard whereby the soundcard has been tested and shown to provide adequate support for applications written around the 'SoundBlaster 16' sound card.
SpeechMike	This is the brand name for the popular PC connected dictation handset from Philips.
Speech Recognition	This is the method of processing sound in the form of spoken words so that a computer can understand commands or create text.
SQL	'Structured Query Language'. SQL is a standard interactive and programming language for getting information from and updating a database. Queries take the form of a command language that allows selection, insertion, updating, location of data, and so forth.
Streaming	Streaming is the process of sending a stream of compressed audio data from a streaming server to a client application that decodes it as it arrives and sends it the sound card for output or writes it to a file on a disk. It is typical that the data will be buffered (See 'Buffering') as playback can be at different speeds to the receipt of streamed audio. A client application that writes captured sound data to a network drive is not regarded as streaming. Only when a recognised streaming service is employed such as that seen with Microsoft server products is it regarded that true streaming is taking place.

T	
Telephony Port	This is a physical connection of a PC to a phone line so that the computer can provide a telephony interface (see 'CTI'). Plug in cards are often used for this and can allow multiple ports per PC.
Terminal Server	The Microsoft Windows Terminal Server (WTS) is a server program that provides the graphical user interface (GUI) of the Windows desktop to user terminals that don't have this capability themselves. The latter include the relatively low-cost "thin client" (see 'Thin-Client') set-up that some companies are purchasing as alternatives to the autonomous and more expensive PC with its own operating system and applications.
Thin Client	A "Thin client" is typically a low-cost, centrally managed computer devoid of CD-ROM players, diskette drives, and expansion slots although it is quite normal to see a desktop PC configured as a thin client as this can deliver administration savings particularly for remotes users. The term 'thin client' derives from the fact that small computers in networks tend to be clients and not servers. Since the idea is to limit the capabilities of these computers to only essential applications, they tend to be purchased and remain "thin" in terms of the client applications they include.

Three Tier

Three-tier is a system architecture consisting of a graphical user interface (GUI) server, an application (business logic or workflow) server, and a database and transaction server. More descriptively, it can be viewed as dividing an application into

- 1) A first-tier, front-end, Web browser-based graphical user interface, usually at a personal computer or workstation.
- 2) A middle-tier business logic application or set of applications, possibly on a local area network or intranet server.
- 3) A third-tier, back-end, database server.

U

V

Vocabulary

This is the term often used to describe a collection of words and the statistics on the use of the words for application in a speech recognition engine (see 'Speech Recognition'). A vocabulary can be a list of general language words or specialist words specific to the user. Many speech recognition applications allow both generic and specialist vocabularies to be used at the same time allowing customisation of the engine to a specific users needs. They also provide for automatic adaptation of these vocabularies as a user uses new words in the normal course of use.

Voice Macro

This is the term used to describe the process within speech recognition software of creating a voice command that can perform one or more actions on a computer.

Voice Recorder

This is term often used to describe a portable digital dictation device.

VOX file

This is a proprietary file type for storing audio data. The most common form of this is the dialogic audio file, which is encoded with the ADPCM codec (see 'ADPCM'). The resulting file has a different structure to the Windows standard wav file (see 'Wav file') so it can not be guaranteed that a VOX file will be readable by any Windows PC without needing proprietary software.

VPN

'Virtual Private Network'. A VPN is a way to use a public telecommunication infrastructure, such as the Internet, to provide remote offices or individual users with secure access to their organization's network. A virtual private network can be contrasted with an expensive system of owned or leased lines that can only be used by one organization. The goal of a VPN is to provide the organization with the same capabilities, but at a much lower cost. VPN is probably the best and most secure method of allowing digital dictation users to work remotely.

W

Wav file

A Wave file is an audio file format, created by Microsoft that has become a standard PC audio file format for everything from system to game sounds. A Wave file is identified by a file name extension of WAV (.wav). Using the adopted standard Resource Interchange File Format (RIFF) it stores amongst other things information such as the audio data, sampling rate (see 'Sample Rate'), data transfer rate (See 'Bytes per Second') plus content or workflow information in recognisable chunks of data within a single file structure. Dictation files stored in .Wav format are almost always guaranteed to be readable by any Windows PC without the need for any proprietary software subject to having the right CODEC (see 'CODEC') installed.

WAN

'Wide Area Network'. A wide area network is a geographically dispersed telecommunications network usually serviced by a telecom provider. The term distinguishes a broader telecommunication structure from a local area network (see 'LAN').

Workflow

Workflow is a term used to describe the tasks, procedural steps, organizations or people involved, required input and output information, and tools needed for each step in a business process. In terms of digital dictation it is the process of using pre defined rules to take speech from multiple authors and turn it into typed documents using a desired typing resource as efficiently as possible.

x

y

z