1

PSTN BASIC INFRASTRUCTURE, INTERFACES, AND SIGNALS

Telephones, fax, and dial-up modems are popularly used with a two-wire TIP-RING foreign exchange subscriber (FXS) interface that supplies a battery. The interface inside the phone or fax machine is a foreign exchange office (FXO). In some countries, a four-wire integrated services digital network (ISDN) is used for telephone services. A T1/E1 family of interfaces is used mainly for higher channel communication. In an office environment, a user may get telephone service through a public switched telephone network (PSTN) central office (CO) or a private branch exchange (PBX) system resident close to the office phones. PBX systems may use multiple FXS or digital phone interfaces for connecting to the user and FXO, ISDN, or T1/E1 family of interfaces to communicate with the nearest PSTN CO or digital loop carrier (DLC). A DLC resides close to the subscribers and extends the reach of central offices. For inter-regional services, the local CO will route the calls to the destination CO. The destination CO then terminates the call directly or through the local DLC. Several handbooks and documents are available on this subject [Freeman (1996), Bellamy (1991), ITU-Handbook (1992)]. The combinations and possibilities in service vary with each service provider and country. VoIP service and user interfaces are closely related to historical PSTN services. In this chapter, an overview of the PSTN telephone infrastructure and some of interfaces and voice signal characteristics is provided to create continuity and to map the PSTN functionality with VoIP infrastructure.

VoIP Voice and Fax Signal Processing, by Sivannarayana Nagireddi

1.1 PSTN CO AND DLC

In this section, an overview of central offices and DLCs is presented. An analog CO provides services to the user, and in recent times, analog central offices were replaced with a combination of digital CO and DLC.

1.1.1 Analog CO

Analog telephony requires a two-wire analog TIP-RING interface. Several years back, the PSTN CO was directly providing several pairs of analog lines to the closest junction box, and from there, individual TIP-RING wire pairs were being distributed to the subscriber [URL (IEC-DLC)]. When a subscriber is far (5000 to 15,000 feet) from the CO, long-distance analog lines can create distortions and signal attenuation. To counter this problem, bigger diameter wires and compensating loading coils were used with analog CO. Sometimes the line voltage is increased at the CO to cater for voltage drop over long lines. For a coverage area of a few miles in diameter, this approach may still be used. Overall, using long lines from the CO is a costly effort, deployment may not scale up, and voice quality degrades. Additionally, fax and modem calls may operate at a lower speed on long analog lines. For a growing customer base, an analog CO may not scale up properly. A VoIP adapter providing telephone service to the end user closely resembles the analog CO by providing telephone interfaces like PSTN DLC or CO and by allowing voice calls through an Internet connection.

1.1.2 Digital CO and DLC

In recent years, PSTN systems migrated to ITU-T-G.711 A-law/μ-law compression [ITU-T-G.711 (1988)]-based synchronous digital communication. Long-distance analog line pairs were replaced with digital distribution boxes. A PSTN CO will connect a few digital lines (T1/E1 family interfaces) to the intermediate DLC box positioned close to the subscriber, and the DLC will then distribute analog TIP-RING lines for the last mile of telephone service. Digital CO along with DLC is shown in Fig. 1.1. With the introduction of the CO and DLC combination, the signal quality is greatly improved, even for the users located far from the CO. Each DLC will take care of several hundred users, and if necessary, the DLC is capable of sending the required high enough voltage to overcome any remaining distance problems. The other side of the DLC is connected to the CO. DLCs may have a basic T1/E1 family of interfaces at the first level. Advanced services, including the support of several offices, PBX systems, multimedia, and Internet capabilities are supported through next-generation DLC (NGDLC) as given in [URL (IEC-DLC)]. NGDLC will need wider bandwidth for communication between the DLC and the CO with fiber being the most popular interface for deployment. In some locations, fiber may not available, leading to a choice of other interfaces,

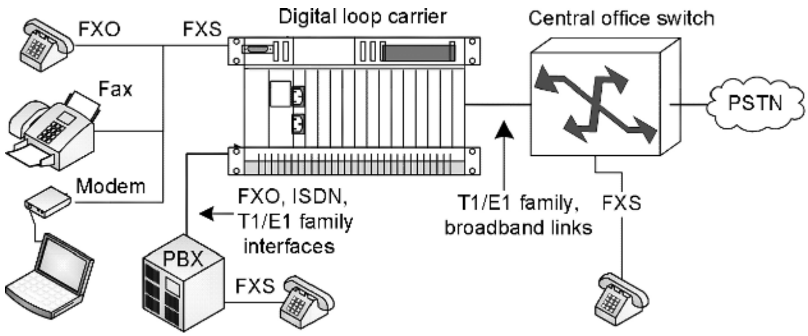

Figure 1.1. PSTN digital office and interfaces.

including coaxial cable, digital subscriber line (DSL), or a combination of these that are used for higher end requirements with NGDLC. NGDLC are deployed for multiservices of voice, video and data as well as various Internet services with the right interfaces.

1.2 PSTN USER INTERFACES

PSTN end users will get services through FXS, ISDN, and the T1/E1 family of interfaces [URL (TIA-496B), URL (T1/E1), URL (ISDN)]. These interfaces are also used in the migration of VoIP voice and fax solutions. As shown in Fig. 1.1, the telephone, fax, and dial-up modem are connected on an FXS interface. The FXO interface is part of a telephone, fax, and modem. Some COs provide an ISDN interface for residential applications. The T1/E1 families of interfaces are mainly used with PBX and enterprise services.

1.2.1 FXS and FXO Analog Interfaces

A PSTN wall socket is the FXS interface given to the subscriber for connecting a telephone. FXS is the two-wire TIP-RING interface provided by the PSTN CO or DLC. This interface is used for connecting telephones, fax machines, and dial-up modems. FXS supplies battery voltage, high-voltage ring, and sufficient current to drive three to five parallel phones. A subscriber line interface circuit (SLIC) and a subscriber line access circuit (SLAC) are the main components of the FXS interface. SLIC consists of a two-to-four-wire hybrid and of high-voltage electronics. SLAC is the interface between a SLIC analog signal and processor digital interface.

The FXO receives battery voltages from the PSTN FXS interface. Sometimes the FXO interface is known as a passive interface, which means the FXO will not generate a high-voltage battery on analog TIP-RING interfaces. The FXO interface is available on the TIP-RING connections from a phone, fax machine, or modem. Subscribers can connect this FXO interface to the FXS

Table 1.1. FXS and FXO Basic Differences

Attribute	FXS	FXO
Main functional category	This works like a central office	This works like an electronic phone
Location in the system	PSTN CO/DLC, PBX, or VoIP adapter gives this interface	Telephones, fax, and dial-up modem are having FXO interface
Where is it connected?	Connected to telephones, faxes, or modems	Central office, PBX, VoIP adapter
Active/passive port	Active port, gives battery voltages of the order of −24 to −72V, usually of −48V	Passive, can receive voltages
Drive capability	Can drive 3 to 5 telephones, or fax machine/modem	Passive interface, present inside phone, fax, or modem
Popular name of the device used on the interface	SLIC is the front end device on DLC or VoIP adapter that drives phones	Passive circuit and DAA is inside the phone, fax, or modem

interface provided by the central office, PBX, or VoIP adapter. PBX systems will use FXS on one side to connect to multiple phones and may use multiple FXO interfaces to connect to PSTN CO or DLC. The other popular name used for this FXO interface is the digital access arrangement (DAA). The DAA name is mainly associated with dial-up modems. In simplest form, multiple FXOs can be connected on one FXS interface. FXS is the main active battery source to supply current to multiple FXO interfaces. FXS and FXO are simple interfaces. Some of the functional differences of FXS and FXO are listed in the Table 1.1.

1.2.2 SLAC, CODEC and codec–Clarifications on Naming Conventions

As explained in Section 1.2.1, telephone interfaces use SLIC and SLAC. SLIC converts a signal from two-wire TIP-RING interface to a four-wire interface and gives it to SLAC for sampling. Currently, the name hardware CODEC is more popular than SLAC. SLAC/CODEC has several functions as explained in Chapter 17. CODEC in a simple configuration consists of a COder [hardware with an analog-to-digit converter (ADC)], and DECoder [hardware with digital-to-analog converter (DAC)] that samples the analog signal at 8- or 16-kHz. FXO interfaces use DAA that also makes use of ADC and DAC for sampling. The name SLIC-SLAC use UPPER-case alphabets. The usage of SLAC equivalent part “CODEC” with SLIC-SLAC and SLIC-CODEC is more common.

The name "codec" with lower-case alphabets is popular for compression. The codec has "encoder for compression" and "decoder for decompression" of digital samples. The names voice codec, speech codec, compression codec, and low-bit-rate codec or simply with lower case "codec" denotes samples compression through computation. Through out this book, the same naming conventions are used.

No hard rules exist for this naming convention. In the literature, lower case name "codec" or "Codec" is also considered for hardware sampling operations, whereas upper-case "CODEC" name is not used to represent compression. Refer the context–sampling hardware (CODEC) or samples compression (codec) for correct interpretation.

1.2.3 TIP-RING, Off-Hook, On-Hook, and POTS Clarifications

The terms TIP, RING, off-hook, and on-hook are used frequently in PSTN and VoIP services. A summary of these keywords is given here. TIP and RING are the pair of wires used for telephone connection. These names originated in reference to the phone plug used to make connections in manual switchboards. In manual switchboards, one side of the line makes contact with the TIP of the plug and the other with the RING contact immediately behind the TIP of the plug [URL (POTS)]. Sleeve is usually of a common point or ground, and this is not used in the current two-wire TIP-RING-based system. Currently, all the telephone exchanges are electronic with digital switching. Even in operator-assisted PBX systems, an operator is using telephone keys or soft keys for making the connection. Normal telephones are not now using TIP, RING, or sleeve connectors. Some instruments for telephone measurements [URL (Sage935)] are using TIP-RING sleeve connectors. The names "TIP, RING" continued for the two wires used for connecting telephones and central offices. A picture to represent this TIP-RING connector is available at reference [URL (POTS)].

The terms "off-hook" and "on-hook" were used in the early days of telephone usage. They refer to the handset position with respect to the cradle of the telephone. In the early telephones, the phone handset was lifted from the hook and was called off-hook. Keeping the back of phone on the hook was called on-hook. On-hook disconnects the phone connection. Off-hook is for getting the dial tone or for continuing the voice call. This hook switch can be clearly seen on most public phones mounted vertically. On public phones, the handset is physically hanging on the hook switch. Lifting the handset from the hook switch is off the hook, and putting the handset back on the hook is the on the hook. A hook switch is also available in handset-based modern phones. The hook may be a little button that pops up on lifting the handset and goes down when putting back the handset.

POTS is a plain old telephone service that was using TIP-RING for main signals, a rotary wheel for pulse dialing, and a bell for ringing. The widely used analog phones of today are also using two-wire TIP-RING with few differences in dial pad, ringer, and connectors. The functions of TIP-RING, dialing, ring, and so on, have remained same. Hence, the current analog phones are also referred as POTS phones. The names "analog phone", "normal phone", "house phone", "POTS phone", "two-wire phone", or simply "phone" are used interchangeably. In this book, the names "ring" or "Ring" are used for phone ringing alert, and name "RING" is used for telephone physical interface wire.

1.2.4 ISDN Interface

ISDN [URL (ISDN)] gives the highest possible PSTN quality than analog telephony. ISDN is used as a reference for comparing voice call quality of narrowband (300 to 3400 Hz) voice. It is also used for high-quality voice and fax at the highest speed of 33,600 bits per second (bps) and for dial-up modem at higher data rates of 56 kilo bits per second (kbps) from one of the two bearer (B)-channels. ISDN is a four-wire interface with a separate send and receive pair of wires. The physical interface of ISDN is of analog, but the transmission content is a bit stream of information. This interface in a system allows simultaneous voice conversations and data communication. With ISDN, voice and data are carried by two B-channels occupying a bandwidth of 64 kbps per channel. Some PSTN switches may limit B-channels to 56 kbps. A data (D) channel handles signaling and data at 16 kbps. There are two types of ISDN services: basic rate interface (BRI) and primary rate interface (PRI). BRI consists of 2B + D, two 64 kbps B-channels and one 16 kbps D-channel for a total of 144 kbps. The basic service is intended to meet the needs of individual users. For enterprise use, PRI is used, and this goes by the name H-series with a suffix based on the channel density. H-channels provide a way to aggregate several B-channels. They are classified as H_0 = 384 kbps (six B-channels), H_{10} = 1472 kbps (23 B-channels), H_{11} = 1536 kbps (24 B-channels), and H_{12} = 1920 kbps (30 useful B-channels). H_{12} is also known as international E1.

1.2.5 T1/E1 Family Digital Interface

T1/E1 is a four-wire interface [URL (T1/E1)]. It uses two wires for send and two wires for receive. T1 is used in North America, and E1 is used in Europe with a T1 equivalent interface named J1 used in Japan. In most of the systems, the same infrastructure and interfaces will work for both T1 and E1 without changing hardware electronics. These interfaces are used for enterprise solutions for voice and data communications. The physical connector interface will look similar to the Ethernet interface. Sometimes two coaxial cables are also used for this interface. PSTN DLC/NGDLC or CO will be supplying this interface to the user. Offices also use the T1/E1 interface for Internet service. In general, T1/E1 is used for voice, data, or a combination of voice and data.

At PSTN CO or DLC, the voice signal coming over the telephone line is sampled at 8 kHz, digitized, and compressed into eight bits producing 64 kbps per line or channel. This per-channel rate of 64 kbps is defined on digital signal 0 (DS0). In T1, 24 DS0 channels are multiplexed into a single digital stream.

The resulting data stream is of $24 \times 8 \times 8$ kHz = 1536 kbps making each frame data 192 bits (1536 kbps/8000). In T1, one-bit synchronization is used for every frame at 8 kHz making each frame data 193 bits and total bit rate 1.544 Mbps. Of 24 channels of T1, one channel is usually allocated for telephone signaling. This process is represented as 23B + D channels, meaning one D-channel is used for data or signaling. In E1, 32 digitized channels are multiplexed into a single digital data stream, resulting in 32×8 bits $\times$ 8 kHz = 2048 kbps. The resulting TDM frame size is 256 bits and frames are at 8 kHz.

1.3 DATA SERVICES ON TELEPHONE LINES

Telephone, fax, and modem use voice band frequencies (300 to 3400 Hz) on the telephone wires. Fax and modem are data terminals. Fax is an independent device that sends data modulated into voice band. A dial-up modem also modulates data into voice band. A dial-up modem is used for Internet connectivity and data transfer. Dial-up modems provide data connectivity up to 56 kbps. Some more details on dial-up modems are given in Chapter 15. Telephones, fax machines, and dial-up modems use telephone lines one at a time, but they can work simultaneously with a broadband service like xDSL.

For VoIP, Internet connectivity is essential. A dial-up modem is one of the basic and most popular options for Internet connectivity. It can provide data connectivity of up to 56 kbps, which is sufficient only with a higher compression VoIP voice call. DSL can provide much higher bandwidth compared with the dial-up modems, and hence, it is used in several deployments for Internet connectivity. In theory, both dial-up and DSL can be used together, but when DSL service is used, a dial-up modem may not be required.

1.3.1 DSL Basics

DSL service is provided through the existing analog telephone two-wire interface or on four-wire ISDN lines. Voice, fax, and modem signals occupy a 0- to 4-kHz band. Telephone wires are capable of carrying wider bandwidth signals to a distance of several thousand feet. DSL technology is based on discrete multi-tone (DMT) modulation using orthogonal frequency division multiplexing (OFDM). It operates on an array of N relatively lower rate transceivers in parallel to achieve an overall high rate on a single line. These N information streams are kept separated from one another by sending them on N different subcarriers called sub-bands or subchannels. Thus, DMT is a very flexible modulation scheme in which each subchannel can carry a different number of bits depending on the channel conditions. The subchannel that supports a higher signal-to-noise ratio (SNR) is loaded with a larger number of bits [Goralski (1998)].

The asymmetric digital subscriber line (ADSL) given in recommendation ITU-T-G.992.1 (1999) specifies DMT modulation with a total of 256 subchannels centered at frequencies $(m)(f_s)$, where m is a subchannel from 0 to 255 and f_s is 4312.5 Hz. Thus, it occupies a bandwidth of (256) (4312.5) = 1.104 MHz on the telephone line. Not all of these subchannels can be used in practice, as voice band splitting filters need to be employed to separate the DSL band from that of the voice band. The recommendation specified for either frequency separation of the downstream and upstream channels or separation by echo cancellation. The maximum number of bits allowed on each subchannel is 15, thereby limiting the theoretical maximum downstream rate that can be achieved to (15 bits per subchannel) (255 subchannels) (4000 symbol rate) = 15.3 Mbps. The G.992.1 recommendation specifies the downstream rate of 8 Mbps and the upstream rate of 1 Mbps.

The asymmetric DSL2 plus (ADSL2+) family [ITU-T-G.992.3 (2005), G.992.5 (2005)] occupies analog signal bandwidth from 25 kHz to 2.2 MHz to support downstream data rates up to 24 Mbps. Very high-speed DSL (VDSL) [ITU-T-G.993.1 (2004), G.993.2 (2006)] will occupy analog bandwidth up to 17 MHz and some versions up to 30 MHz to support data rates of 50 to 100 Mbps and more. In summary, voice services are continued at low frequency while using the same telephone line for various flavors of DSL service.

The DSL access multiplexer (DSLAM) is the CO for DSL services. DSL services do not need a high-voltage battery for operation. Hence, isolation in combining and splitting is essential between high voltage for normal telephone operation and low voltage for wideband DSL signals. DSL has to be combined on analog signal lines only. Different combining techniques are used to provide DSL service. Some popular techniques of combining DSL are direct analog combining, remote DSLAM line based, ADSL lines cards, and remote access multiplex (RAM) [URL (IEC-ADSL)]. Overview on DSL combining methods are given below. It is expected that these techniques and technologies will keep growing and changing with time.

Direct Analog Combining. This technique is applicable for the analog COs that distribute direct analog lines. DSL signals are directly combined on the analog lines with isolation splitter-combiners. This type of operation is possible if the coverage distance is less and the bandwidth requirements are of a few Mbps. Most central offices extend services through DLCs; hence, this type of DSL support is not popular.

Remote DSLAM-Based Combining. In this scheme, a complete DSLAM is colocated along with DLCs or is positioned close to the DLC based on the available installation options. DSLAMs are usually maintained in a controlled environment and are managed remotely. The service can be used with any PSTN lines and several flavors of DSL. The main disadvantage of this method is the cost involved to deploy and maintain. It will also require long wiring if DLC and DSLAM are not located in the same racks.

ADSL Line Card Solutions. DLC will have multiple slots. DSLAM in line cards are arranged in either a separate rack similar to DLC or some of the

DLC box slots be reserved for the DSL line cards. Keeping the DSL cards in DLC may minimize the wiring requirements. Management of DSL and DSLAM line cards will be under the same framework. Many deployments use this approach. DSL flavors keep changing with time. Upgrading the line cards in the field is a major concern of this deployment, and multiple user cards may go through different interoperability issues.

RAM. RAM performs similar functions of providing DSL service. RAM integrates closely with the DLCs. It is a separate small box. The RAM boxes are also called pizza boxes based on their shape. Pizza boxes are stacked and positioned along with the DLCs. RAM-based boxes have the advantage of both remote DSLAM and ADSL line card solutions. These boxes are easy to upgrade, change with another version of box, and scale in density.

1.4 POWER LEVELS AND DIGITAL QUANTIZATION FOR G.711 μ/A-LAW

Telephone line electrical signal power levels are presented in dBm units. Power level expressed in dBm is the power in decibel (dB) with reference to one milliwatt (mW). Power 0 dBm represents one milliwatt of power. One mW is a strong speech level for listening. Usual telephone conversion levels are close to –16 dBm (25 μW; μW is microwatt). ITU-T-G.711 and ITU-T-G.168 (2004) recommendations listed speech power levels and corresponding digital quantization mapping for 8-bit A-law and μ-law. A-law and μ-law (pronounced as Mu) are logarithmic compression methods explained in Chapter 3 under G.711 compression. Telephone interfaces have SLIC and SLAC for converting TIP-RING signals to digital samples. SLIC converts a signal from two-wire TIP-RING interface to a four-wire interface to SLAC. In recent literature, the hardware CODEC name is used instead of SLAC. CODEC consists of a COder, DECoder hardware with an analog-to-digital converter (ADC) and a digital-to-analog converter (DAC) that samples the signals at 8 or 16 kHz. Sampled signals are given to the processor for additional processing. In this chapter, signal levels and the mapping to 16-bit numbers in A-law and μ-law, are given with reference to dB and dBm scale.

1.4.1 μ-Law Power Levels and Quantization

The maximum μ (read as Mu)-law undistorted power level for the sine wave is 3.17 dBm. Power of 3.17 dBm is 2.0749 mW in 600 Ohms (Ω) of telephone impedance. Impedance of 600 Ω is applicable to North America and a few other countries like Japan and Korea. A sine wave of 2.0749 mW in 600 Ω develops 1578 milliVolts (mV) peak or a 3.156 Volts peak-to-peak sine wave amplitude. Voltage of 1578 mV has to be quantized as 8159 amplitude [ITU-

T-G.711 (1988), ITU-T-G.168 (2004)] in 16-bit binary numbering format. Assuming linear mapping for calculation purposes, the quantization level is 1578 mV/8159 = 0.1934 mV. Once voice samples are available as per the above quantization of one level per 0.1934 mV, the following formula given in ITU-T-G.168 is used for estimating the voice power over a block of samples:

$$S(k) = 3.17 + 20\log\left[\frac{\sqrt{\frac{2}{n}\sum_{i=k}^{k-n+1} e_i^2}}{8159}\right] \quad \mu\text{-law encoding} \tag{1.1}$$

$S(k)$ = signal level in dBm
e_i = linear equivalent of the pulse codec modulation (PCM) encoded signal at sample index i
k = discrete time index or starting index for block of samples
n = block size of samples or number of samples over which root mean square (RMS) measurement is made.

In general, N will be used in signal processing literature for representing blocks of samples and n will be used for an index within the block of samples. In the above equation, to maintain consistency with the ITU recommendations, n is used for denoting block size and i is used for the index.

A sine wave with a peak amplitude of 8159 results in 3.17 dBm without any dependency on frequency. µ-law coding distorts the sine wave if it exceeds 3.17 dBm. Square wave power is 3 dB more than sine wave power, and an amplitude of 8159 of square wave results in 6.17 dBm. Alternatively, a constant value of 8159 (constant voltage) is 3 dB more power than a sine wave of 8159. Sometimes use of a constant value and calibrating with reference to a constant of 8159 as 6.17 dBm can help for calibration of the algorithms and power estimation blocks. In practice, lower amplitude is used for such calibrations in the range of –13 to –10 dBm instead of 6.17 dBm. Coding with µ-law supports the SNR up to 37 to 41 dB (for a 0- to –30-dBm signal power level) for the useful speech signal levels.

1.4.2 A-Law Power Levels and Quantization

An A-law maximum undistorted power level for the sine wave is 3.14 dBm. Power of 3.14 dBm is 2.06 mW in 600 Ω of telephone impedance. A sine wave of 2.06 mW in 600 Ω will develop 1572 mV peak or a 3.144 Volts peak-to-peak sine wave amplitude. Voltage of 1572 mV has to be quantized to 4096 amplitude levels. Quantization level is 1572 mV/4096 = 0.3838 mV. Once voice samples are available as per the above quantization of one level per 0.3838 mV, the following formula given in ITU-T G.168 (2004) is used for estimating the voice power over a block of samples:

$$S(k) = 3.14 + 20\log\left[\frac{\sqrt{\frac{2}{n}\sum_{i=k}^{k-n+1} e_i^2}}{4096}\right] \quad \text{A-law encoding} \tag{1.2}$$

The symbols used in this equation are the same as those used in the μ-law power representation in Eq. (1.1). For a sine wave of peak 4096, this calculation results in 3.14 dBm with out any dependency on frequency. A-law starts distortion to sine wave if it exceeds 3.14 dBm. An amplitude of 4096 for the square wave results in 6.14 dBm. As an example, substituting one complete cycle of a 2-kHz sine wave at 4096 amplitude (four samples for one complete cycle are 0, 4096, 0, –4096) in A-law power calculation gives 3.14 + 0 dBm = 3.14 dBm. Substituting four constant values of 4096 in A-law power calculation will give 3.14 + 3.0 dBm = 6.14 dBm. A-law supports SNR up to 37 to 41 dB (this varies with signal power level) for the useful speech signal levels.

1.5 SIGNIFICANCE OF POWER LEVELS ON LISTENING

Table 1.2 is given for different power levels starting from +3.17 dBm and down to –66 dBm in a few suitable steps. In relation to Table 1.2, summary points are given here [ITU-Handbook (1992)]. These levels are mainly mapped to the telephone interfaces and voice chain processing built for VoIP. In VoIP, power levels play a role in operation of some voice processing operations such as voice activity detection (VAD)/comfort noise generation (CNG), automatic gain control (AGC), echo level, and ideal channel noise. PSTN systems do not have any power-sensitive module operations such as VAD/CNG. In voltage scale, a 0.775-Volt RMS sine wave is the same as 0 dBm, terminated in 600-Ω impedance. The usual speech conversation level is –16 dBm in power. Some members may speak at a low amplitude below –24 dBm, which can also happen because of positioning of acoustic interfaces during conversation. This problem demands more effort by the listener. Automatic gain control (AGC), if employed, starts working from this power level. As per the ITU-T-G.169 (1999) recommendation, a maximum gain of 15 dB is used on low-level speech signals. A peak-to-average (PAR) level of speech is 15 to 18 dB [ITU-Handbook (1992)] for 0.1% of the active speech time. An average speech level of –12 dBm with 15 dB peak to average can preserve the speech as undistorted.

In practice, speech signals are gain controlled up to speech levels of –42 to –40 dBm. Below –42 dBm, it is difficult to continue the communication with required perception of speech. Lower power levels below –42 dBm are mapped to the region of VAD/CNG zones in VoIP. In PSTN, this type of silence region and power distinction is not present. VAD/CNG algorithms given in Chapter 4 track up to –60 dBm of power. Note that VAD/CNG algorithms use units of

Table 1.2. Summary on Useful Power Levels and Corresponding A- and μ-Law Mapping Values

Power in dBm	Power	Sine wave peak voltage in 600 Ω (milliVolts)	A-law peak 16-bit number	μ-law peak 16-bit number	Remarks on power levels
3.17	2.0749 mW	1578	Distortion starts	8159	μ-law undistorted sine wave
3.14	2.0606 mW	1572	4096	8131	A-law undistorted sine wave
0	1.0000 mW	1095	2853	5664	Loud speech
–3	0.5012 mW	776	2020	4010	
–6	0.2512 mW	549	1430	2839	
–9	0.1259 mW	389	1012	2010	
–12	63.1 μW	275	717	1423	Usual conversation level is –16 dBm
–15	31.6 μW	195	507	1007	
–18	15.8 μW	138	359	713	
–20	10.0 μW	110	285	566	
–21	7.9 μW	97.6	254	505	
–24	4.0 μW	69.1	180	357	Low-level speech, AGC action starts from –24 dBm
–27	2.0 μW	48.9	127	253	
–30	1.0 μW	34.6	90	179	
–33	0.5 μW	24.5	64	127	
–36	0.25 μW	17.4	45	90	
–39	0.125 μW	12.3	32	64	
–42	62.5 nW	8.7	23	45	AGC action ends, VAD action may start
–50	10.0 nW	3.5	9	18	Instruments maintain reasonable accuracy up to –52 dBm
–60	1.0 nW	1.1	3	6	Long delayed echoes are perceived as audible
–66	0.25 nW	0.5	1	3	Close to ideal channel noise of –68 dBm (–68 dBm = 20 dBrnc)

Abbreviations: mW, milli Watt; μW, micro Watt; nW, nano Watt

dBov, and the suffix ov stands for overload. It is not related to units of voltage, and it is related to 6.17 dBm = 0 dBov in μ-law compression. More details on these aspects are given in Chapter 4.

When round-trip delay is more, echoes are clearly perceivable. Echo levels given in Chapter 6 are suppressed to lower than –65 dBm. At this lower power level of –65 dBm, one or two bits in a 16-bit numbering system will appear as listed in Table 1.2. An echo power level of even –60 dBm is considered disturbing in VoIP and inter-regional PSTN calls. When no voice is speaking into the telephone or the phone is in mute, some noise overrides in the telephone system because of front-end coupling and power supplies. This noise is called ideal channel noise. These noise levels are maintained to below –68 dBm and preferably to –80 dBm. Noise power is expressed in picoWatts (1 picoWatt = 10^{-12} Watt) on a dB scale as dBrnC; suffix "C" denotes the C-message filter. A noise level of 90 dBrn is equivalent to 1 mW or 0 dBm [Bellamy (1991)].

1.6 TR-57, IEEE-743, AND TIA STANDARDS OVERVIEW

In PSTN, communication from the originating CO or DLC to the destination DLC is synchronous digital communication. This process has been well established over the last several decades. PSTN transmission follows some common standards such as ITU and IEEE and country-specific standards like TR-57 and TIA standards [TR-NWT-000057 (1993), TIA/EIA-470C (2003), IEEE-STD-743 (1995)]. PSTN manufacturers use TR-57 in North America to specify the signaling and transmission requirements for DLC. The FXS interface on VoIP adapters works similar to DLC or analog CO providing the TIP-RING interface. Assuming other favorable conditions for VoIP voice quality, as explained in Chapter 20, ensuring TR-57 transmission characteristics helps in improving voice quality. TR-57 is used in North America, and the current standard is GR-57 [GR-57-CORE (2001)]. Several other standards are followed in different countries. For example, Japan makes use of the NTT PSTN specification [URL (NTT-E)], and France uses the France Telecom STI series [FT ITS-1 (2007)] of specifications. Most of these country specifications are closely related to each other with only a few deviations like end-to-end losses, transmission filter characteristics, impedance, and so on. TR-57 transmission specifications can also be followed in achieving voice quality in VoIP systems. In VoIP, some signaling events as call progress tones used in establishing end-to-end calls may go through longer delays than PSTN specification. These deviations will tend to develop during the call establishment phase. Meeting transmission characteristics [TIA/EIA-912 (2002)] can take care of established VoIP call voice quality.

1.6.1 TR-57 Transmission Tests

TR-57 or GR-57 transmission summary points and high-level interpretation are given in this section. [Courtesy: Text and parameter values with TR-57 reference are from "TR-NWT-000057 (1993), Functional criteria for digital loop carrier systems—Chapter 6"; used with the permission from Telcordia, NJ, www.telcordia.com]. It is suggested to refer to TR-57 [TR-NWT-000057 (1993)] or the recent document GR-57 [GR-57-CORE (2001)] for the complete requirements and detailed specifications referred to in this book. Integrating these test specifications into PSTN and VoIP helps in maintaining the quality of voice and fax calls. Refer to instrument manuals [URL (Sage-options)] for details on measurements and interpretation.

Return Loss. Return loss happens because of an impedance mismatch of the terminations and the source. It is expressed as the ratio of the outgoing signal to the reflected signal. It is given as echo return loss (ERL) and singing return loss (SRL) with SRL at both high (Hi) and low (Lo) frequencies. ERL is the frequency-weighted average of the return losses over the 3-dB bandwidth points of 560 to 1965 Hz. Higher ERL is advantageous and gives lower echo. ERL in dB has to be greater than 18 dB, and this level avoids echo issues in the intraregional calls. In numerical terms, 18-dB ERL means 1/64th of the power or 1/8th of the voltage returning as echo. It is possible to achieve ERL up to the extent of 24 dB in practical systems with short-loop applications with VoIP gateways. To clarify on a relative scale, an ERL of 24 dB is better than an ERL of 18 dB by 6 dB. SRL is created by a resonance effect at selected frequencies. At some frequencies, the round trip of the echo is created with multiples of 360° of phase shift, which makes oscillations sustain for a longer duration. This phenomenon is called "line is on singing." SRL-Lo is the singing return loss for 3-dB bandwidth low frequencies in the range of 260 to 500 Hz. SRL-Hi is for the high-frequency singing return loss measured in the band of 2200 to 3400 Hz. Details on the band-pass characteristics are given in IEEE STD-743 [(1995), URL (Midcom)]. The SRL has to be greater than 10 dB. In general, SRL-Hi will be higher than SRL-Lo. Lower dB values of SRL-Hi, SRL-Lo creates annoyance to both talker and listener of the voice call.

Longitudinal Balance. Balance is the match between TIP and RING lines. It provides the discrimination for the common mode signal on the TIP and RING interface or indicates how well the TIP-RING is matching. It is evaluated as the ratio of longitudinal to metallic balance. Longitudinal is a common mode signal, and metallic is a differential signal. In the measurements, a common or ground point is required in addition to TIP-RING lines, as given in reference [URL (Sage-Balance)]. The preferred balance is greater than 65 dB at 1 kHz, but the acceptable balance is 63 dB. In good implementations, a balance of 68 dB is achieved.

Total Loss. It identifies that the transmitter and receiver are having a required loss. The end-to-end losses have to be within the acceptable limits, including the network cables loss. PSTN systems achieve their highest end-to-end loss of 8 dB in off-hook, and these losses vary with country-specific requirements. These losses also influence the loudness rating (given in Chapters 17 and 20) that has a close relationship with voice quality.

Off-Hook Frequency Response. Frequency response has to be flat to meet set tolerances for most of the voice band. This test also ensures that frequency response roll off did not happen in the 400- to 2800-Hz band. Frequency response is checked at 0 dBm, and the flatness goals are of ±0.5 dB. The reference for comparison of flatness is at 1004 Hz. With reference to a 1004-Hz frequency response, deviations from 400 to 2800 Hz have to be within ±0.5 dB as optional and −0.5 to +1.0 as required. In actual practice, the frequency response is taken from 100 Hz to 3900 Hz in suitable frequency steps. The values between 400 and 2800 Hz are used for comparing the frequency response.

60-Hz Signal Loss. Signal loss at 60 Hz is not desirable in speech transmission, but this "mains hum" can unfortunately be introduced from the AC mains power supply (110 V, 60 Hz). This test is to ensure that a 60-Hz contribution is at least 20 dB lower than a frequency response at 1004 Hz. This 60-Hz contribution has to be treated before sampling. If end-to-end loss at 1004 Hz is "X" dB, loss at 60 Hz has to be greater than X + 20 dB. In some countries, especially in Europe, 50 Hz is used for power transmission, and therefore, the measurements have to be conducted at 50 Hz in those countries even though it is not stated.

Off-Hook Amplitude Tracking. It ensures transmission gain flatness for different amplitudes at 1004 Hz, which is similar to amplitude linearity at the selected frequency. Three different ranges are defined in TR-57. At lower signal levels, A-law and μ-law create more quantization noise, which makes gain deviation to increase below the −37-dBm input level. For inputs of greater than or equal to −37 dBm, maximum deviation is ±0.5 dB (average of ±0.25 dB). For inputs of −37 to −50 dBm, maximum deviation is ±1.0 dB. For inputs of −50 to −55 dBm, maximum deviation is ±3.0 dB, which implies that AGC cannot improve linearity for small signals. Most instruments can measure up to a minimum of −60 dBm. Test instruments use a C-message band-pass filter [URL (Sage935)] for the voice band. In some countries, psophometric (p)-message filters are used. The differences are outlined in the ideal channel noise section.

Overload Compression Tests. The A-law can accept 3.14 dBm and the μ-law can accept 3.17 dBm of a sine wave without distortion. With increased signal power, the sine wave is clipped on either end of the top to make it look like a trapezoidal waveform. Overload compression is for measuring distortions on overloaded power levels up to 9 dBm. These distortions are measured as loss in simple measurements. There are set limits to the maximum loss at different power levels of 0, +3, +6, and +9 dBm. If system loss is "X" dB at 0-dBm input, loss at +3 dBm has to be ≤X + 0.5. The loss at +6 dBm has to be ≤X + 1.8. The loss at the highest signal of +9 dBm has to be lower than ≤X + 4.5. The relations are good for no loss of X = 0 dB. When loss is significant, the measurements may show much better results. Assuming loss at an analog interface, even a higher power signal may pass without any distortions.

Off-Hook Ideal Channel Noise. In its simplest form, ideal channel noise is the noise heard when microphones are in mute. The noise can be heard in a silent room. Instruments in North America use a C-message filter for voice band and measure the noise power in dBrnc. Power in dBrnc (also referred as dBrnC) is C-message weighted noise power with a reference unit of 1 pw. The popular notation is dBrnc for noise power. For voice power, dBm units are used. In simple mapping, 90 dBrn = 0 dBm (1 mw), and with C-message weighting, the mapping is 88 dBrnc = 0 dBm. The psophometric weighting is used in Europe as 87.5 dBp = 0 dBm. The symbol "dBp" is used for dBrn with psophometric weighting. The variation is by 2 to 2.5 dB with the weighting windows. The ideal channel noise power has to be better than 18 dBrnc, which is equivalent to 18 – 88 = –70 dBm. An accepted noise power level is 20 dBrnc = –68 dBm. Long-distance analog telephony was accepting higher ideal channel noise [Bellamy (1991)]. After conversion to digital telephony with short analog loops and DLCs, ideal channel noise achieved better levels, as stated in this section.

SNR or Distortion Ratio. The SNR and distortion measure identifies the SNR limits at the signal level. Lower signal power levels and their SNR results are also important in transmission. A-law or μ-law compression is used in PSTN and VoIP. Both the highest and the lowest signal are coded in 4 bits with 3-bit level scaling and in 1 bit for sign (polarity of the signal), which limits the possible SNR to 38–41 dB with self-quantization effects. With low amplitudes, SNR starts degrading. It is measured with an input sine wave of 1004 to 1020 Hz, and signal power is varied from 0 to –45 dBm. For input of 0 to –30 dBm, the required SNR is more than 33 dB (usually >37 dB is achieved in most systems). For input of –30 to –40 dBm, SNR is >27 dB. For –40 to –45 dBm, SNR is >22 dB. When end-to-end losses are more, or ideal channel noise is not meeting the specifications, SNR will be lower for the same input signal levels. The signal-to-total-distortion (STD) ratio that takes into account total distortion, including second and third harmonics.

Impulse Noise. Impulse noise is the short duration noise power that exceeds the set thresholds. Impulse noise can create audible tick sounds, and fax/modems may fall back to a lower bit rate. Ideal channel noise is the steady low-level noise. Impulse noise durations are comparable with 10-ms duration. Signal loss from phase or amplitude distortion can be noticed during VoIP packet loss, coupling from adjacent channels, ringing, line reversals clock drifts, and a change in power supply load conditions in the hardware. Noise on the interfaces may occupy wider bandwidth. Impulse noise is heard with a C-message filter, which means it is limited to the voice band.

The number of impulse counts in 15 minutes should not exceed 15; therefore, one impulse count per minute is the maximum. The measurements are conducted on both ends of the system with tones and simple ideal channel configuration. In the no-tone case, impulses are detected with a threshold of

47 dBrnc (equivalent to –41 dBm). In the tone case, with the tone at 1004 Hz, –13 dBm is sent and impulse noise is measured with a threshold of 65 dBrnc (equal to –23 dBm). Several conditions with impulse noise exist as given in TR-57 [TR-NWT-000057 (1993)]. The impulse noise counts are also referred to as hits.

Intermodulation Distortion. This test makes use of four tones as per IEEE STD-743 (1995) with a total input level of –13 dBm. This test is mainly used for measuring the undesired nonlinearity. A pair of tones around 860 Hz and another pair of tones around 1380 Hz are used in this test. In this test, total second and third harmonic distortions are measured. The second harmonic has to be better than 43 dB, and the third harmonic has to be better than 44 dB.

Single-Frequency Distortion. It measures the distortions to the single pure tone. This distortion is separated into two frequency bands of 0–4 kHz and 0–12 kHz. In narrowband, the 1004–1020 Hz tone at 0 dBm is sent and the observed tones outside the sent tone have to be lower than –40 dBm in the frequency band of 0–4 kHz. In the wideband range, for any input at 0 dBm in the frequency range of 0–12 kHz, the observed output at any other frequency outside the sent tone frequency has to be less than –28 dBm in the 0–12-kHz frequency range. This type of test will require spectrum analyzers.

System Generated Tones. It measures any undesired system generated tones. Send and receive can be in either the terminated mode or the not terminated mode with proper impedance. Tones measured at any end have to be less than –50 dBm in the frequency range of 0–16 kHz.

PAR. PAR provides amplitude and phase distortion over time because of transmission impairments [URL (Sage-PAR)]. A PAR waveform is a complex signal consisting of about 16 non-harmonically related tones with a spectrum that approximates the modem signals in the voice frequency band. In general, PAR takes care of attenuation distortion, envelope distortion, four-tone modulation distortion, and phase distortion. PAR is measured on both ends. Most systems achieve a PAR of 94, and an acceptable PAR is 90. The PAR measurement cannot reveal what caused this distortion. In many types of equipment, PAR test is replaced with the better 23-tone test as per IEEE STD-743 (1995).

Clarification Note on PAR Measurement and PAR Speech Levels. In specifying the speech characteristics, peak-to-average levels are specified as 15.8–18 dB. The usual speech levels are of –16 dBm, and peak levels can go up to +2 dBm with an 18-dB speech-level PAR. The PAR specified in TR-57 and the measurements using instruments from [URL (Sage-PAR)] provide the distortion measure for voice band signals. A PAR of 100 is the best, and a PAR of 90 to 94 is the goal in digital-loop carrier [TR-NWT-000057 (1993)]. A PAR of 94 is 6% of overall transmission distortions.

Channel Crosstalk. This test is to ensure, end-to-end, that no significant coupling occurs in the adjacent channels. Therefore, 0-dBm power tones from 200 to 3400 Hz are sent and power is measured in adjacent channels. This crosstalk power has to be 65 dB lower or below −65 dBm of power. With this type of low coupling, crosstalk will submerge into ideal channel noise.

Frequency Offset. It measures the frequency drift from end to end, which happens mainly because of a sampling clock parts per million (PPM) difference between source and destination. The frequency drift has to be contained to within 0 ± 0.4 Hz at 1004 Hz. The instruments have to cater to a 0.1-Hz measurement resolution.

1.6.2 IEEE STD-743–Based Tests

IEEE STD-743 (1995) had many tests that overlapped with TR-57 measurements. Instrument manufacturers were following IEEE specifications. Several tests of TR-57 are contained in IEEE STD-743. The 23-tone test has been adapted in recent test equipment. It takes care of transmission and distortion tests. The 23-tone test signal consists of 23 equally spaced tones [URL (Sage935)] from 203.125 to 3640.625 Hz with known phase relationships. The tones are spaced at 8000/512 = 15.625 Hz to allow easy fast Fourier transform (FFT) analysis with 512 and 256 points at 8000-Hz sampling. The signal repeats once every 64 ms (1/15.625 Hz = 64 ms). Frequency response is measured at 23 tones, and phase relations are observed at 22 frequencies.

1.6.3 Summary on Association of TR-57, IEEE, and TIA Standards

Many equipment manufacturers built instruments as per 1984 version of IEEE STD-743. This outlines the IEEE standard accuracy of various measurements and measurement procedures. It has many overlapping measurements of TR-57. IEEE STD-743–based instruments can perform TR-57 tests also. IEEE STD-743 specifications were revised in 1995. This revision had many tests suitable for TR-57. This standard became obsolete in 2004. Equipment manufacturers, however, still follow IEEE STD-743. Now the TIA-470 series is followed in many types of new equipment. TIA-470 has had several revisions and extensions for acoustic measurements, such as cordless phones and wireless systems that were not available in TR-57- and IEEE STD-743–based tests. TR-57 is limited to a narrowband of 0–4 kHz. IEEE STD-743 (1995) and TIA IEIA-470C (2003) take care of wideband voice tests. In general, all these standards have several matching and overlapping specifications that determine proper voice quality on PSTN-based systems and their extensions across many different types of telephone and fax equipment.