Maintaining Audio Quality in the Broadcast and Netcast Facility

2019 Edition

Robert Orban
Greg Ogonowski

orban
AUDIO PROCESSING

HIFI AUDIO
STREAMS
ENCODERS

DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet
ailsDetailsDetailsDetailsDetailsDetailsDetailsDetai
lsDetailsDetailsDetailsDetailsDetailsDetailsDetails
DetailsDetailsDetailsDetailsDetailsDetailsDetailsD
etailsDetailsDetailsDetailsDetailsDetailsDetailsDe
tailsDetailsDetailsDetailsDetailsDetailsDetailsDet

Maintaining Audio Quality in the Broadcast and Netcast Facility

2019 Edition

Robert Orban
Greg Ogonowski

orban
AUDIO PROCESSING

HIFI AUDIO
STREAMS
ENCODERS

Published 25 February 2019.

Rorb Inc., Belmont CA 94002 USA

Modulation Index LLC, 1249 S. Diamond Bar Blvd Suite 314, Diamond Bar, CA 91765-4122 USA
Phone: +1 909 860 6760; E-Mail: info@indexcom.com; Site: https://www.indexcom.com

Table of Contents

Maintaining Audio Quality in the Broadcast/Netcast Facility

By Robert Orban (Orban) and Greg Ogonowski (StreamS)

Authors' Note

In 1999, we combined and revised two previous Orban publications on maintaining audio quality in the FM and AM plants and have since revised the resulting publication several times. In 2019, considerations for both AM and FM are essentially identical except at the transmitter because, with modern equipment, there is seldom reason to relax studio quality in AM plants. The text emphasizes FM, HD Radio, and netcasting practice; differences applicable to AM have been edited into the main text.

Preface

This publication is organized into four main parts and an appendix:

1. **Recording media:** Compact disc, CD-R and CR-RW, DVD±R, DVD±RW, DVD-A, HD-DVD, Blu-ray, digital tape, magnetic disk, flash RAM, and data compression are discussed in the body of this document (see page 5), while vinyl disk, phonograph equipment selection and maintenance, analog tape, tape recorder maintenance, recording alignment tapes and cart machine maintenance are discussed in *Appendix: Analog Media* starting on page 81.

2. **System considerations:** headroom, audio metering, measuring loudness, voice/music balance, and electronic quality—see page 23.

3. **Configuring and Using the production studio:** choosing monitor loudspeakers, loudspeaker location and room acoustics, loudspeaker equalization, production software, other production equipment, and production practices—see page 49.

4. **Equipment following the audio processor:** netcast encoders, FM exciters, transmitters, and antennas—see page 75.

5. **Appendix: Analog Media**: Mostly material from older editions of this document, with some light editing. It is mainly of historical interest, although the section on vinyl playback may still be relevant. See page 81.

NOTE: Because the state of the art in audio technology is constantly advancing, it is important to know that this material was last revised in late 2018. Our comments and recommendations obviously cannot take into account later developments. We have tried to anticipate technological trends when that seemed useful.

There is an enormous amount of information available on the Internet. A considerable amount of it is incorrect and can often mislead you. Be very careful about where you obtain information outside of this guide.

Introduction

Achieving excellent on-air and on-line quality audio is more important than ever because of the intensely competitive nature of available media. "Good enough" no longer works. Persuading audiences to listen and then return requires attention to quality at every stage of the production process, from program sources to audio processing.

This guide will help you achieve your quality on-air and on-line audio broadcasting and netcasting goals.

Achieving quality audio is difficult. It requires serious commitment on several levels. You might think that in this digital age, software would make this trivial. But software varies widely in quality and. will only be as good as its operator, assuming the software works correctly to begin with. Knowledge is crucial.

Achieving a BIG sound requires financial investment, but this kind of sound creates revenue opportunities and easily pays for itself. You must avoid the "rinky-dink" tinny radio sound that has become synonymous with so much on-air and on-line content. This is a "turn-off" in every sense of the words and will drive listeners away.

Your audio source content *must* be pristine. It all starts at the source and many audio source problems *cannot* be fixed with realtime processing. Your on-air and on-line audio content needs professional audio processing to achieve consistency and level control, as well as accurate, high-performance radio transmitters and audio encoders to preserve the integrity of the controlled audio. Your software and hardware tools must be chosen very carefully; their performance is *not* equivalent. You can't simply "play records" into a transmitter and/or encoder and hope that can compete. You won't!

The same, if not better, efforts that are used for your terrestrial on-air sound quality should now be applied to your streaming audio content. Many terrestrial radio stations go through painstaking efforts to deliver audio quality that attracts and holds listeners. This includes starting with good audio source material, using capable professional audio playout software, and applying professional audio processing to provide a consistent, polished, "branded" sound to their audience. Streaming has now become just as important because more and more audience is listening to streams instead of traditional terrestrial transmitters. This also means that you should use a professional, high-performance streaming audio encoder to deliver the streams to your audience. Not all streaming audio encoders sound equally good at a given bitrate. They need to be chosen wisely.

Almost all new automobiles are now equipped with Apple CarPlay and Google Auto. It is literally easier to connect to a stream from the dashboard than to tune the radio. Moreover, the stream has the added potential of greater coverage with much better audio quality. With StreamS efficient AAC Encoders, bandwidth usage and cost are no longer a restriction.

Reliability is crucial. Your transmission system should be provisioned with adequate fail-over protection.

The “Digital Divide”

Broadcasting and netcasting now rely heavily on Information Technology (IT). Achieving consistent state of the art audio quality in broadcast is a challenging task. It begins with a professional attitude, considerable skill, patience, and an unshakable belief that quality is well worth having. It usually requires the careful cooperation between programming, engineering, and computer IT departments. With the advent of computer-based audio systems and computer network-delivered audio, it requires extra computer IT knowledge. Computer IT personnel should understand digital audio fundamentals. Broadcast engineers and IT professionals tend to have different skillsets. Contemporary broadcasting and netcasting require merging these two so that practitioners are competent in both. There should be no “digital divide.”

The “digital divide” refers to IT specialists and broadcast audio specialists who are experts in their own specialties but may lack understanding about others. This typically results in poor communication, misunderstandings, and suboptimal technical solutions to problems that require input from both areas of expertise. Broadcast and computer industries have different terminology, with many same or similar terms having different meanings. This “lingo” needs to be understood correctly by both departments.

Just because something is possible doesn’t mean it’s advisable. Many multimedia developers have done more harm than good to the broadcast and netcast industry by poor development and lack of understanding of the business, resulting in poor software and hardware solutions. Not embracing existing standards and protocols has been another problem, which causes system interface incompatibilities. This moves things backward, not forward.

To best serve audiences, digital and computer technologies are supposed to move multimedia forward, and the digital divide needs to be bridged for this to happen optimally: IT specialists need to learn about professional audio, both analog and digital, and broadcast/audio specialists need to learn about networking and streaming. In this spirit, we hope that this book will be useful not only to broadcast/audio specialists, but also to IT specialists.

This document provides some technical insights and tips on how to achieve immaculate audio, and keep it that way. Remember, successful broadcasting and streaming *all start at the source*.

Audio Processing: The Final Polish

Audio processors change certain characteristics of the original program material in the quest for positive benefits such as increased loudness, improved consistency, and absolute peak control to prevent distortion in the following signal path and/or to comply with government regulations.

The art of audio processing is based on the idea that such benefits can be achieved while giving the listener the illusion that nothing has been changed. Successful audio processing performs the desired electrical modifications while presenting a subjective result that sounds natural and realistic. This sounds impossible, but it is not.

Audio processing provides a few benefits that are often unappreciated by the radio or television listener. For example, the reduction of dynamic range caused by processing makes listening in noisy environments (particularly the car) much less difficult. In music having a wide dynamic range, soft passages are often lost completely in the presence of background noise. Few listeners listen in a perfectly quiet environment. If the volume is turned up, subsequent louder passages can be uncomfortably loud. In the automobile, dynamic range cannot exceed 20 dB without causing these problems. Competent audio processing can reduce the dynamic range of the program without introducing objectionable side effects.

Further, broadcast program material typically comes from a rapidly changing variety of sources, most of which were produced with no regard for the spectral balances of others. Multiband limiting, when used properly, can automatically make segues between sources much more consistent. Multiband limiting and consistency are vital to the station that wants to develop a characteristic audio signature and strong positive personality, just as feature films are produced to maintain a consistent look. Ultimately, it is all about the listener experience.

Good broadcast operators are hard to find, making artful automatic gain control essential for the correction of errors caused by distractions or lack of skill. Also, the regulatory authorities in most countries have little tolerance for excessive modulation, making peak limiting mandatory for signals destined for the regulated public airwaves.

Optimod-FM, Optimod-AM, Optimod-DAB, Optimod-TV, and Optimod-PCn have been designed to address the special problems and needs of broadcasters and netcasters while delivering a quality product that most listeners consider highly pleasing. However, every electronic communication medium has technical limits that must be fully heeded if the most pleasing results are to be presented to the audience. For instance, the audio quality delivered by Optimod is highly influenced by the quality of the audio presented to it. If the input audio is very clean, the signal after processing will probably sound excellent, even after heavy processing. Distortion of any kind in the input signal is likely to be exaggerated by processing and, if severe, can end up sounding offensive and unlistenable.

Audio processing is an art and the “sound” of a given audio processor is a function of hundreds of variables, many of which involve trade secrets known only to their manufacturers. (This includes Orban.) Comparing audio processors by counting the number of bands of compression/limiting or listing other features obvious from the front panel is superficial and futile. Processors must be judged on how they perform with the many different types of program material used in a given format and ultimately should be judged based on their ability to attract and hold a given broadcaster's target audience. There is no substitute for long-term listening.

AM/MW is limited by poor signal-to-noise ratio and by limited receiver audio bandwidth (typically 2-3 kHz). As delivered to the consumer, it can never be truly "high fidelity." Consequently, multiband audio processing for AM compresses dynamic range more severely than in typical FM or digital practice. In addition, pre-emphasis (whether NRSC or more extreme than NRSC) is required to ensure reasonably crisp, intelligible sound from typical AM/MW radios. In AM, this is always provided in the audio processor and never in the transmitter.

Audio quality in TV viewing is usually limited by small speakers in the receivers, although widespread adoption of DTV, HDTV, personal entertainment, and home theatre has changed some of this, increasing consumer demand for high audio quality. In everyday television viewing, it is important to avoid listener irritation by maintaining consistent subjective loudness from source to source. A CBS Loudness Controller combined with multiband processing, both included in Optimod-TV and Optimod-PCn, can achieve this.

Netcasting (also known as webcasting), DAB, and HD Radio almost always use low bitrate codecs. Processing for such codecs should not use clippers for limiting, and should instead use a look-ahead type limiter (see page 79). Optimod-Surround, Optimod-DAB, Optimod-FM (HD processing chain), and Optimod-PCn provide the correct form of peak limiting for these applications and other low bite rate digital audio services.

Just as the motion picture industry creates a consistent, professional look to their product by applying exposure and color correction to every scene in a movie, audio processing should be used as part of the audio broadcast product to give it that final professional polish.

Part 1: Recording Media

Compact Disc

The compact disc (CD) is an important source for recorded music for broadcasting. With 16-bit resolution, 44.1 kHz sample rate, and no lossy compression (unlike MP3), it provides excellent source quality for radio. Although digital downloads are now available with up to 24-bit resolution and sample rates up to 192 KHz, these are unlikely to provide audible advantages in broadcasting/netcasting, and the increasing presence of digital watermarks makes them potentially a lower-quality source than older, un-watermarked CDs. DVD-Audio, which also offers 24-bit resolution and 96 kHz sample rate, was a commercially unsuccessful and now obsolete attempt to raise the quality of consumer media, while SACD, which uses “bitstream” coding (DSD Direct-Stream Digital) instead of the CD’s PCM (Pulse Code Modulation), is as of this writing hanging on as a niche format, primarily for classical music. Most prognosticators believe that the future of audiophile-quality music lies in downloadable high-resolution files that use lossless compression. These files are already available from several sources.

Because most audio is still sourced at a 44.1 kHz sample rate, upsampling to 48 kHz does not improve audio quality. Further, many broadcast digital sources have received various forms of lossy data compression, like MP3.

Although digital-to-analog conversion technology is constantly improving, we believe that some general observations could be useful. In attempting to reproduce CDs with the highest possible quality in the analog domain, the industry has settled into technology using “delta-sigma” digital-to-analog converters (DACs) with extreme oversampling. These converters use pulse width modulation or pulse-duration modulation techniques to achieve high accuracy. Instead of being dependent on the precise switching of voltages or currents to achieve accurate conversion, contemporary designs depend on precise timing, which is far easier to achieve in production.

Oversampling simultaneously increases the theoretical signal-to-noise ratio and produces (prior to the reconstruction filter within the CD player) a signal that has no significant out-of-band power near the audio range. A simple, phase-linear analog filter can readily remove this power, ensuring the most accurate phase response through the system. We recommend that CD players used in broadcast employ technology of at least this quality when connected to the broadcast facility via analog connections. However, the engineer should be aware that these units might emit substantial amounts of supersonic noise, so the low-pass filtering in the transmission audio processor must be sufficient to reject this to prevent aliasing in digital transmission processors or STLs.

The broadcast environment demands ruggedness, reliability, and quick cueing from audio source equipment. CD players intended for live, on-air use must also be chosen for their ability to track even dirty or scratched CDs with minimum audible artifacts, and on their ability to resist external vibration. There are dramatic differences between players in these areas! We suggest careful comparative tests between players using imperfect CDs to determine which players click, mute, skip, or otherwise mistrack. Striking the top and sides of the player with varying degrees of force while listening to the output can give a “feel” for the player’s vibration resistance. Fortunately, some of the players with the best sound also track best. The depressing trade-off between quality and ruggedness that is inevitable in vinyl disk reproduction is unnecessary when CDs are used.

Reliability is not easy to assess without experience. The experience of your fellow broadcasters can be valuable here—ask around during local broadcast engineers’ meetings. Be skeptical if examination of the “insides” of the machine reveals evidence of poor construction.

Cueing and interface to the rest of the station are uniquely important in broadcast. There are, at this writing, relatively few players that are specifically designed for broadcast use—players that can be cued by ear to the start of a desired selection, paused, and then started by a contact closure. The practical operation of the CD player in your studio should be carefully considered. Relatively few listeners will notice the finest sound, but all listeners will notice miscues, dead air, and other obvious embarrassments!

A design that tries to minimize CD damage caused by careless handling places each CD in a protective plastic “caddy.” The importance of handling CDs with care and keeping the playing surface clean cannot be over-emphasized. Contrary to initial marketing claims of invulnerability, CDs have proven to require handling comparable to that used with vinyl disks in order to avoid broadcast disasters.

Except for those few CD players specifically designed for professional applications, CD players usually have unbalanced –10dBV outputs. In many cases, it is possible to interface such outputs directly to the console (by trimming input gains) without RFI or ground loop problems. To solve any problems, several manufacturers produce low-cost–10dBV to +4dBu adapters for raising the output level of a CD player to professional standard levels. However, a digital connection (via AES3 or SPDIF) from the player to the console will always work better.

CD-R and CD-RW, DVD±R, DVD±RW, DVD-A, HD DVD, Blu-ray

Recordable optical media are attractive as audio sources and for archiving. They have error detection and correction built in, so when they working correctly, their outputs are bit-for-bit identical to their inputs. Recordable CD, DVD, and Blu-ray discs are available.

There are several dye formulations available and manufacturers disagree on their archival life. However, it has been extrapolated that any competently manufactured CD-R should last at least 30 years if it is stored at moderate temperatures (below 24° C / 75° F) and away from very bright light like sunlight. On the other hand, these disks can literally be destroyed in a few hours if they are left in a locked automobile, exposed to direct sunlight. The industry has less experience with more recent formats like DVD-R and Blu-ray. No recordable optical medium should be considered to be archival without careful testing.

Archiving CD-R in data format is better than archiving in Red Book audio format because the error correction in data format is more robust.

Not all media of a given type are equal. Choose media to minimize bit-error-rate (BER). At the time of this writing, Taiyo Yuden, TDK, and Verbatim are known to have low BER. However, manufacturers will change formulations and plants from time to time, so these recommendations may not be valid in the future.

The reflectivity of a good CD-R is at best 90% of a mass-produced aluminized CD. Most CD players can accommodate this without difficulty, although some very old players cannot. Because of the lower reflectivity, the lasers within broadcast audio CD players need to be in good condition to read CD-R without errors. Sometimes, all that is necessary is a simple cleaning of the lens to restore satisfactory performance.

CD-RW (compact disk–rewritable) is not a true random-access medium. You cannot randomly erase cuts and replace them because the cuts have to be unfragmented and sequential. However, you can erase blocks of cuts, always starting backwards with the last one previously recorded. You can then re-record over the space you have freed up.

The disadvantage of CD-RW is that some CD payers cannot read them, unlike CD-R, which can be read by almost any conventional CD player if the disk has been "finalized" to record a final Table of Contents track on it. A finalized CD-R looks to any CD player like an ordinary CD. Once a CD-R has been finalized, no further material can be added to it even if the disk is not full. If a CD-R has not been finalized, it can only be played in a CD-R recorder, or in certain CD players that specifically support the playing of unfinalized CD-Rs.

HDCD

Originated by Pacific Microsonics and later purchased by Microsoft[1] when Pacific Microsonics folded, HDCD is a method of encoding a wider dynamic range than 16-bits into a standard audio CD. It uses two techniques to do this is a way that is reasonably compatible with non-decoded playback: Peak Extend, which is a reversible soft limiter; and Low Level Range Extend, which is a reversible gain on low-level signals. Low Level Range Extend provides a benefit at the expense of a very minor increase in noise in non-decoded playback. However, Peak Extend can cause audible distortion in non-decoded playback.

Additionally, the HDCD process dynamically switches the characteristics of the encoder's anti-aliasing filter and the player's reconstruction filter to complement the program material on a moment-to-moment basis, purportedly improving transient response while minimizing audible aliasing.

In 2001, there were approximately 5000 titles available with HDCD decoding and if you use one of these as a source, it is wise to decode the material and store it linear PCM form with a resolution of at least 24 bits. This is necessary because the HDCD decoder will reduce most peak levels by 6 dB and will also reduce low levels. If Peak

---

[1] Microsoft discontinued the official HDCD website in 2005, but since version 9 of Windows Media Player, it has been possible to decode and play HDCD-enabled on Windows computers using a 24-bit soundcard and WMP. Note that Windows Media Player will *not* rip a decoded HDCD. The resulting 16-bit PCM file will be HDCD encoded.

Extend was used, the decoder will exploit the extra 6 dB of dynamic range to restore peak levels applied to Peak Extend's soft limiter. However, a significant number of releases use little or no Peak Extension.

There are several Windows utilities that claim to decode HDCD files into linear PCM, although none incorporate the filter switching. HDCD.exe is a command-line tool that requires no installation. It is probably quickest and most convenient to use, and is scriptable for batch processing. There is an HDCD Decoder plug-in available for the foobar2000 application[2]. It offers realtime HDCD status indication, but requires installation and careful configuration.

It is also possible to record the analog output of an HDCD-equipped hardware CD player using 24-bit/96 KHz capture and gain substantially all of the benefits of the HDCD process (including the filtering) if you use a high-quality A/D converter for the capture and take great care to avoid adding hum, noise, and clipping in the analog path. See *Analog Connections* on page 64.

M-DISC

M-DISC[3] sells archival media using a stone-like formulation that, according to the company, provide a 1000-year archival lifetime. These are available in DVD and Blu-ray formats. They can be read by any DVD or Blu-ray reader but DVDs must be burned with M-DISC compatible writers, which are available from several major manufacturers. A list of compatible writers is available on the M-DISC website. M-DISC Blu-rays can be burned in any conventional Blu-ray writer.

Digital Tape

While DAT was originally designed as a consumer format, it achieved substantial penetration into the broadcast environment. This 16-bit, 48 kHz format is theoretically capable of slightly higher quality than CD because of the higher sample rate. In the DAR environment, where 48 kHz-sample rate is typical, this improvement can be passed to the consumer. However, because the "sample rate" of the FM stereo system is 38 kHz, there is no benefit to the higher sampling rate by the time the sound is aired on FM.

The usual broadcast requirements for ruggedness, reliability, and quick cueing apply to most digital tape applications, and these requirements proved to be quite difficult to meet in practice. The DAT format packs information on the tape far more tightly than do analog formats. This produces a proportional decrease in the durability of the data. To complicate matters, complete muting of the signal, rather than a momentary loss of level or high frequency content, as in the case of analog, accompanies a major digital dropout.

---

[2] https://www.foobar2000.org/components/view/foo_hdcd

[3] http://www.mdisc.com/

At this writing, there is still debate over the reliability and longevity of the tape. Some testers have reported deterioration after as little as 10 passes, while others have demonstrated almost 1000 passes without problems. Each demonstration of a tape surviving hundreds of passes shows that it is physically possible for R-DAT to be reliable and durable. Nevertheless, we now advise broadcasters not to trust the reliability of DAT tape for long-term storage and never to use it for new recordings. Always make a backup, particularly because DAT is now an obsolete format and finding players in working order is more and more difficult. If your facility has DAT tapes in storage, it would be wise to copy them to other media as soon as practical.

Hard Disk Systems

Hard disk systems use sealed Winchester hard magnetic disks or optical disks (originally developed for mass storage in data processing) to store digitized audio. This technology has become increasingly popular as a delivery system for material to be aired. There are many manufacturers offering systems combining proprietary software with a bit of proprietary hardware and a great deal of off-the-shelf hardware. If they are correctly administered and maintained, these systems are the best way to ensure high, consistent source quality in the broadcast facility because once a source is copied onto a hard drive, playout is consistent. There are no random cueing variations and the medium does not suffer from the same casual wear and tear as CDs. Of course, hard drives fail catastrophically from time to time, but RAID arrays can make a system immune to almost any such fault. (However, it is still wise to back up a RAID array; they can fail catastrophically and lose data.)

It is beyond the scope of this document to discuss the mechanics of digital delivery systems, which relate more to ergonomics and reliability than to audio quality. However, two crucial issues are how the audio is input and output from the system, and whether the audio data is stored in uncompressed (linear PCM) form or using some sort of data compression.

Audio is usually input and output from these systems through sound cards. Please see the discussion on page 34 regarding sound cards and line-up levels.

Flash RAM

As its price continues to fall, flash RAM has become ever more popular as an audio storage medium. It is available, packaged with a controller, with many different interfaces, USB2.0 being one of the most popular and universally compatible. Unlike CD-RW and DVD-RW, flash RAM is capable of tens of thousands of writes and hundreds of thousands of reads. It can be written to and read from in faster than real time for any commonly used audio sample rate and bit depth.

Flash RAM is available in two main technologies, called NOR and NAND-types. Both are usable for audio recording and have different trade-offs. While neither is capable of true random access writes because old data must be block-erased, this is usually not a limitation for audio recording.

The long-term storage reliability of flash RAM has not yet been proven and it is therefore unwise to rely on flash RAM as a sole means for archival backup.

This article provides a good summary of flash RAM technology:
http://en.wikipedia.org/wiki/Flash_memory

Program Associated Data (PAD) or Metadata

Metadata is data describing data. Audio metadata can be included in audio files, where it is commonly called *tags*, and/or included in the playout system database of professional playout software. Metadata can also be used to transport Now Playing information. Audio metadata is a commonly misunderstood topic, even by software developers and providers.

Although not directly affecting audio quality, metadata is discussed here because it is crucial to understand how it affects workflow. The exact details are beyond the scope of this document, although the basics are presented. Metadata and transport protocols used to move this data can be complicated.

Perhaps the most common example of metadata is the ID3 tag format, found in MP3 files. There are now several versions of ID3 tag formats, some of which are not backward-compatible because of the advanced features that have been added. ID3v2.4 is preferred as of this writing, as it supports UTF-8 character encoding, which supports all character sets. The ID3 tag format has also become a standards-based tagging format now used in many other audio file formats beyond MP3. It is extensible, providing the ability to add customized specific tags. Other common tagging protocols include Broadcast Wave Format (BWF), RIFF WAVE Info Chunk, ISO MP4 Box iTunes Metadata, and FLAC tags. Compliant HLS Audio-Only streams use ID3v2.4 for metadata.

There is a myth that .wav files cannot be tagged or contain metadata. This is *not* true; it is part of the extensible RIFF specification used for .wav files. For example, any CD ripped using Microsoft Windows Media Player to .wav format is tagged using a RIFF ListInfo Chunk

There are many ways to implement audio metadata, which can be standards-based or proprietary. It is important to understand these differences, as it has everything to do with your production workflow, playout software, and Now Playing information for RDS, HD Radio, and streaming services. Metadata and transport protocols are also used for remote control GPIO, content insertion (such as ad replacement), and target loudness control.

For these applications, it is important to use transports and streaming protocols that have low latency. Not all streaming protocols can provide low-latency metadata, and this is one of many important considerations for using HLS to deliver audio streams.

Using incorrect protocols here are responsible for awkward-sounding ad replacements. The StreamS HLSdirect™ Encoders deliver low latency, on time metadata.

Playout system software usually supports at least one metadata format. This can be used to populate the database entries upon audio ingest. For proper importation, it is important to use the metadata format that your playout system software supports. Some playout software does not require metadata tagged files for import. Instead, it may be possible to import the metadata as a text file with audio file references. RadioDJ Pro is one such example. This conveniently facilitates pre or post editing metadata in one place and doesn't require audio files to be tagged, making it very convenient to import high-quality untagged .wav files into playout software. The database entries are then used to display information about the audio tracks in the playout software and also to send Now Playing data to other destinations in your facilities.

It is common for playout system vendors to use proprietary metadata in their audio files to try to lock their users into their systems, making it difficult to use audio libraries in another system. There are audio file and metadata utilities that can fix this problem, and in our opinion the playout system providers have wasted resources developing inconvenient proprietary protocols that provide no real protection for them. The same is true of many streaming service providers. Beware of this practice.

In addition to the actual metadata, certain services such as RDS, HD Radio, and streaming have an associated transport layer to get the metadata from source to destination, usually over a computer network using either TCP/IP or UDP/IP. There are many details to the transport layer as well. TCP/IP and UDP/IP have different modes and directions. For a successful source and destination metadata connection, all of these details must be understood. There are bridge software applications available such as StreamS PADbridge™ to make this easier, and also to add many additional metadata features to improve the user experience.

Data Compression

Data compression is ubiquitous, and choosing the correct compression algorithm (codec) for delivery to the consumer is crucial to maintaining audio quality. Almost all digital audio is delivered via some form of data compression algorithm. Hence, digital playout systems that use data compression should use the highest quality codec possible because the audio will be compressed again at transmission. Cascaded codecs can cause severe and unexpected loss of audio quality.

There are two forms of compression—lossy, and lossless. Best modern practice is to use lossless or no data compression in an audio playout system; this will yield significant audible benefits.

Lossless Compression

Lossless compression provides an output that is bit-for-bit identical to its input. The only standards-based lossless codec is MPEG-4 ALS (formerly LPAC). This has provisions for tagging and metadata. Some other lossless codecs include Windows Media Lossless (used in Windows Media Player), Apple Lossless (used in QuickTime and iTunes), FLAC (Free Lossless Audio Codec), WavPack, and Shorten. WinZip 11.0 and above uses WavPack to compress .wav files and writes them to .zipx format.

All of these algorithms remove statistical redundancy in the audio signal to achieve approximately 2:1 compression of audio that has not been heavily processed. They have lower coding efficiency with material that has been subject to heavy dynamics compression and peak limiting, like much of today's music.

Because lossless audio codecs are transparent, their usability can be assessed by speed of compression and decompression, compression efficiency, robustness, error correction, file tagging features, and software and hardware compatibility. Unless there is an error or bug in the implementation of the codec, it is almost impossible for different lossless codecs to sound different. Although one could conceive of a scenario where the different algorithms load a decoding computer's CPU differently and hence introduce different amounts of jitter into an onboard DAC via ground currents or power supply modulation, we are unaware of any evidence that this has ever actually been demonstrated. Unless an audible difference between lossless compression algorithms survives a double-blind listening test, it is safe to assume that any such claims have no physical reality and are caused by the "expectation" or "placebo" effect in the mind of the listener.

Lossy Compression

Lossy compression eliminates data that its designer has determined to be "irrelevant" to human perception, permitting the noise floor to rise instead in a very frequency-dependent way. This exploits the phenomenon of *psychoacoustic masking,* which means that quiet sounds coexisting with louder sounds will sometimes be drowned out by the louder sounds so that the quieter sounds are not heard at all. The closer in frequency a quiet sound is to a loud sound, the more efficiently the louder sound can mask it. There are also "temporal masking" laws having to do with the time relationship between the quieter and louder sounds.

A good psychoacoustic model that predicts whether an existing sound will be masked is complicated. The interested reader is referred to the various papers on perceptual coders that have appeared since the late 1980s in EBU references and in the professional literature, mostly in the *Journal of the Audio Engineering Society* and in various AES Convention preprints.

There are two general classes of lossy compression systems, non-psychoacoustic and psychoacoustic. The first is exemplified by ADPCM and APT-X®, which, while designed with full awareness of psychoacoustic laws, do not contain psychoacoustic models. In exchange for this relative simplicity they have a very short delay time (less than 4ms), which is beneficial for applications requiring foldback monitoring, for example.

The second class contains built-in psychoacoustic models, which the encoder uses to determine what parts of the signal will be thrown away and how much the noise floor can be allowed to rise without its becoming audible. More advanced codecs (like MPEG-2 AAC) contain adaptive filterbanks that minimize audible pre-echo on transients. These codecs can achieve higher subjective quality for a given bitrate than codecs of the first class at the expense of much larger time delays. Examples include the MPEG family of encoders, including Layer 2, Layer 3, AAC, and HE-AAC (also known as aacPlus). The Dolby® AC-2 and AC-3 codecs also fall in this category. The large time delays of these codecs make them unsuitable for any application where they are processing live microphone signals that are then fed back into the announcer's headphones. In these applications, it is sometimes possible to design the system to bypass the codec, feeding the undelayed or less-delayed signal into the headphones.

There are two general applications for codecs in broadcasting — "contribution" and "transmission." A contribution-class codec is used in production. Accordingly, it must have high enough "mask to noise ratio" (that is, the headroom between the actual codec-induced noise level and the just-audible noise level) to allow its output to be processed and/or to be cascaded with other codecs without causing the codec-induced noise to become unmasked and without introducing audible pre-echo.

A transmission-class codec is the final codec used before the listener's receiver. Its main design goal is maximum bandwidth efficiency. Some codecs, like Layer 2, have been used for both applications at different bitrates. There are many proprietary, non-MPEG codecs other than Dolby AC3 available, but these are not standards-based and are beyond the scope of this document.

Ideally, all codecs implementing a given standards-based algorithm (for example MPEG1 Layer 2 or AAC) have equal performance. However, this is not true in practice. Codec standards emphasize standardizing the decoders while allowing the encoders to be improved over time. While it is expected that not all manufacturer's encoders will perform equally, to a less extent this is also true of decoders. Not every decoder realizes the standard in an ideal way—for example, there can be compromises caused by using fixed-point arithmetic in a codec whose reference code was implemented in floating point. There can also be numeric inaccuracies caused by the sample-rate conversion algorithms that are often included in the codec implementation. Not all codecs of the same type have equal performance.

To assess the audible transparency of codecs, the ITU has published Recommendation ITU-R BS.1116-1, which is intended for use in the assessment of systems that introduce impairments so small as to be undetectable without rigorous control of the experimental conditions and appropriate statistical analysis. All of the high-quality MPEG standard-based codecs have been assessed using this algorithm and the results have been published.

Similarly, the ITU has developed the BS.1534-1 standard, commonly known as MUSHRA (MUltiple Stimulus with Hidden Reference and Anchors), which is widely used for the evaluation of systems exhibiting intermediate quality levels, in particular low-bitrate codecs. MPEG standard codecs HE-AACv1 and HE-AACv2 have been tested using this standard and the results have been published.

To our knowledge, there is no published, neutral, third-party work that assesses the Windows Media Audio® codec family using the BS.1116-1 and BS.1534-1 methodologies. Hence, we believe that the MPEG-standard codecs have more credibility.

MPEG1 — Layer 2/3

The MPEG1 layer 2 and layer 3 codecs are the oldest-technology codecs still in general use today. Layer 2 is a sub-band audio encoder, which means that compression takes place in the time domain with a low-delay filter bank producing 32 frequency domain components. By comparison, Layer 3 is a transform audio encoder with hybrid filter bank, which means that compression takes place in the frequency domain after a hybrid (double) transformation from the time domain.[4] Layer 2 and Layer 3 have very different ancestry and different sets of trade-offs. Layer 2 can offer audibly transparent performance at high bitrates and is free from the pre-echo artifacts that plague Layer 3. However, Layer 3's subjective performance degrades less abruptly as bitrate is reduced.

Layer 3, also known as MP3, was designed for use as a transmission codec and is not audibly transparent on all program material regardless of bitrate, although modern implementations can sound very good at rates of 256 kb per second and higher. Over the years, there have been many MP3 encoder implementations having widely varying audio quality. Because of this, it is undesirable to use Layer 3 / MP3 files for playout system sources if high audio quality is an objective.

Moreover, AAC (see below) is at least 30% more bitrate-efficient than MP3, and AAC can be audibly transparent at higher bitrates. Another disadvantage of MP3 is that its stereo coding can add L–R energy, which can increase multipath distortion in FM stereo broadcast. For these reasons, AAC is almost always preferred over MP3 for any application. The only reason to offer an MP3 stream for transmission is that some old hardware players (such as those in old home theater receivers) do not support AAC.

Because Layer 2 can be audibly transparent at high bitrates, it is suitable for both contribution and transmission, although its low coding efficiency compared to AAC (Advanced Audio Coding) make it obsolete for transmission.

Layer 2 audio can be easily identified by viewing using an FFT spectrum analyzer. The upper bands or bins will appear to gate on and off as the audio plays.

AAC/HE-AAC

MPEG-2 Advanced Audio Coding was designed for use as a transmission codec as the successor to MPEG-1 Layer 3. It incorporates numerous improvements to the Layer 3 algorithm[5], including significantly improved pre-echo performance. Blind tests show

---

[4] http://en.wikipedia.org/wiki/MPEG-1_Audio_Layer_II

[5] http://en.wikipedia.org/wiki/Advanced_Audio_Coding

that AAC demonstrates greater sound quality and transparency than MP3 for files coded at the same bitrate.

AAC/HE-AAC codec technology combines three MPEG technologies: AAC, Coding Technologies Spectral Band Replication (SBR), and Parametric Stereo (PS). SBR is a bandwidth extension technique that enables audio codecs to deliver similar quality at half the bitrate of codecs that do not use SBR. Parametric Stereo increases the codec efficiency a second time for low bitrate stereo signals.

SBR and PS are forward and backward compatible methods to enhance the efficiency of any audio codec. AAC was chosen as the core codec for HE-AAC because of its superior performance over older generation audio codecs such as MP3 or WMA. This was one of the main reasons why Apple Computer chose AAC for their market-dominating iTunes downloadable music service.

Members of the HE-AAC Codec Family

HE-AACv1 combines AAC and SBR. **HE-AACv2** builds on the success of HE-AACv1 and adds more value where higher compression efficiency for stereo signals is required. HE-AACv2 is a true superset of HE-AACv1, as HE-AACv1 is of AAC. HE-AACv2 adds Parametric Stereo to HE-AACv1, further improving coding efficiency at low bitrates.

HE-AAC delivers streaming and downloadable audio files at 48 kbps for FM-quality stereo, entertainment-quality stereo at 32 kbps, and good quality for mixed content even below 16 kbps mono. This efficiency makes new applications in the Internet, mobile, and digital broadcast markets viable. Moreover, unlike certain other proprietary codecs, AAC/HE-AAC does not require proprietary servers for streaming. AAC/HE-AAC can be stream-delivered using SHOUTcast, Icecast2, QuickTime/Darwin, Real/Helix, Adobe Flash and Wowza Media Servers.

The members of the HE-AAC codec family are designed for forward and backward compatibility. Besides HE-AACv2 bit streams, an HE-AACv2 encoder is also capable of creating HE-AACv1 and plain AAC bit streams.

Every decoder is able to handle bit streams of any encoder, although a given decoder may not exploit all of the stream's advanced features. An HE-AACv2 decoder can fully exploit any data inside the bit stream, be it plain AAC, HE-AACv1 (AAC+SBR), or HE-AACv2 (AAC+SBR+PS). An AAC decoder decodes the AAC portion of the bit stream, not the SBR portion. As a result, the output of the decoder is bandwidth limited, as the decoder is not able to reconstruct the high frequency range represented in the SBR data portion of the bit stream.

If the bitstream is HE-AACv2, an AAC decoder will decode it as limited-bandwidth mono and an HE-AACv1 decoder will emit a full-bandwidth mono signal; an HE-AACv2 decoder is required to decode the parametric stereo information.

Extended HE-AAC/xHE-AAC[6] (xHE-AAC) is the latest upgrade to the MPEG AAC family. It significantly improves the audio quality of music and speech particularly at very low bitrates of 8 kbit/s to 32 kbit/s, and is compatible with HE-AAC streams. It combines and improves upon HE-AACv2 for music and generic audio and AMR-WB+ for speech. USAC further builds on the technologies in MP3 and AAC and takes these one step further. It includes all the essential components of its predecessors and improves them.

Extended HE-AAC combines the advantages of existing speech and music codecs. By adding a new set of encoding tools to the HE-AACv2 audio codec, Extended HE-AAC outperforms dedicated speech and general audio coding schemes and bridges the gap between both worlds, providing consistent high quality audio for all signal types. Accordingly, Extended HE-AAC can improve the quality of existing low bitrate services or more audio channels can be transmitted at a given bitrate.

xHE-AAC's advantages over AAC-LC/HE-AAC include improved Spectral Band replication (SBR; see page 19) at medium bitrates, improved Parametric Stereo (PS; see page 21) at medium bitrates, improved speech quality at low bitrates, improved signaling and transport for immediate decoding of SBR and PS, scaling to transparency at high bitrates, and preservation of the same AAC/HE-AAC structure. xHE-AAC decoders decode AAC-LC/HE-AACv1/2 bitstreams, which simplifies player development.

USAC preserves the same overall structure of HE-AACv2. The core coder consists of an AAC based transform coder, enhanced by ACELP speech coding technology AMR-WB+. An enhanced Spectral Band Replication (SBR) tool, eSBR, handles high frequencies, while MPEG surround 2-1-2 supplies parametric stereo coding. By combining HE-AACv2 and AMR-WB+ and improving on both, USAC becomes a unified speech and music codec that performs equally well on all types of audio content, at all bitrates.

eSBR extends SBR to cope with low core coder sample frequencies, which are usually used at very low bitrates. Other SBR improvements offer better performance at medium bitrates. Adding Predictive Vector Coding (PVC) to eSBR introduces a new coding scheme for SBR. This improves the subjective quality of the eSBR tool, for both speech and music.

USAC stereo coding is improved by extending, combining, and integrating both discrete and parametric stereo (PS) coding while improving coding of interchannel

---

[6] Neuendorf, Max; Multrus, Markus; Rettelbach, Nikolaus; Fuchs, Guillaume; Robilliard, Julien; Lecomte, Jérémie; Wilde, Stephan; Bayer, Stefan; Disch, Sascha; Helmrich, Christian; Lefebvre, Roch; Gournay, Philippe; Bessette, Bruno; Lapierre, Jimmy; Kjörling, Kristofer; Purnhagen, Heiko; Villemoes, Lars; Oomen, Werner; Schuijers, Erik; Kikuiri, Kei; Chinen, Toru; Norimatsu, Takeshi; Chong, Kok Seng; Oh, Eunmi; Kim, Miyoung; Quackenbush, Schuyler; Grill, Bernhard: "The ISO/MPEG Unified Speech and Audio Coding Standard—Consistent High Quality for All Content Types and at All Bitrates," J. AES Volume 61 Issue 12 pp. 956-977, December 2013.

phase shifts. This results in improved stereo image stability and perceptual sound-field envelopment that more closely matches the source.

Signaling and transport of MPEG-D USAC is similar to MPEG-4 HE-AACv2, but a USAC decoder unambiguously determines its configuration at startup, and there is no delay in SBR or PS. USAC decoders will decode AAC-LC, HE-AACv1, HE-AACv2, and xHE-AAC bitstreams.[7]

Standardization

AAC/HE-AAC is an open standard. While not a proprietary format (unlike other less efficient codecs), at this writing it is not royalty-free to encoder and player manufacturers, but does not require users to pay royalties for streaming or file encoding if they are using a licensed encoder. AAC/HE-AAC is widely standardized by many international standardization bodies, as follows:

- MPEG 2 AAC
- MPEG ISO/IEC 13818-7:2004 Advanced Audio Coding
- MPEG 4 AAC
- MPEG ISO/IEC 14496-3:2001 Coding of Audio-Visual Objects — Audio, including Amd.1:2003 Bandwidth Extension, Amd.2:2004 Parametric Coding for High Quality Audio, and all corrigenda
- MPEG 4 HE-AACv1 = AAC LC + SBR (aka HE-AAC or AAC+)
- MPEG ISO/IEC 14496-3:2001/AMD-1: Bandwidth Extension
- MPEG-4 HE-AACv2 = AAC LC + SBR + PS (aka Enhanced HE-AAC or eAAC+)
- MPEG ISO/IEC 14496-3:2001/AMD-2: Parametric Coding for High Quality Audio
- MPEG Surround [Spatial Audio Coding (SAC)]
- MPEG ISO/IEC 23003-1:2007 Part 1: MPEG Surround
  MPEG ISO/IEC 23003-2:2010 Part 2: Spatial Audio Object Coding (SAOC)
  MPEG ISO/IEC 23003-2:2012 Part 3: MPEG-D USAC

HE-AACv1 is standardized by 3GPP2 (3rd Generation Partnership Project 2), ISMA (Internet Streaming Media Alliance), DVB (Digital Video Broadcasting), the DVD Forum, Digital Radio Mondiale, and many others. HE-AACv2 is specified as the high

---

[7] See https://www.indexcom.com/products/encoder/encodertechnology/ for a comparison of the performance of the member of the MPEG codec family.

quality audio codec in 3GPP (3rd Generation Partnership Project) and all of its components are part of MPEG-4.

As an integral part of MPEG-4 Audio, HE-AAC is ideal for deployment with the H.264/AVC video codec standardized in MPEG-4 Part 10. The DVD Forum has specified HE-AACv1 as the mandatory audio codec for the DVD-Audio Compressed Audio Zone (CA-Zone). Inside DVB-H, HE-AACv2 is specified for the IP-based delivery of content to handheld devices. ARIB has specified HE-AACv1 for digital broadcasting in Japan. S-DMB/MBCo has selected HE-AACv1 as the audio format for satellite multimedia broadcasting in Korea and Japan. Flavors of MPEG-4 HE-AAC or its components or portions thereof are also applied in national and international standards and systems such as iBiquity's HD Radio (US), XM Satellite Radio (US), and or the Enhanced Versatile Disc EVD (China).

Independent quality evaluations of AAC/HE-AAC

Independent tests have clearly demonstrated HE-AACv2's value. In rigorous double-blind listening tests conducted by 3GPP (3rd Generation Partnership Project), HE-AACv2 proved its superiority to its competitors even at bitrates as low as 18 kbps. HE-AACv2 provides extremely stable audio quality over a wide bitrate range, making it the first choice for all application fields in mobile music, digital broadcasting, and the Internet.

HE-AACv1 has been evaluated in multiple 3rd party tests by MPEG, the European Broadcasting Union, and Digital Radio Mondiale. HE-AACv1 outperformed all other codecs in the competition.

The full "EBU subjective listening test on low bitrate audio codecs" study can be downloaded at: http://tech.ebu.ch/publications/tech3296.

In 2018, the best overall quality for a given data rate in a transmission codec is achieved by the MPEG AAC codec (at rates of 96 kbps or higher) and Extended HE-AAC (at rates below 96 kbps). The AAC codec is about 30% more efficient than MPEG1 Layer 3 and about twice as efficient as MPEG1 Layer 2. The AAC codec can achieve "transparency" (that is, listeners cannot audibly distinguish the codec's output from its input in a statistically significant way) at a stereo bitrate of 128 kb/sec, while the Layer 2 codec requires about 256 kb/sec for the same quality. The Layer 3 codec cannot achieve transparency at any bitrate, although its performance at 192 kbps and higher is still very good.

Spectral Band Replication

Low bitrate audio coding is an enabling technology for a number of applications like digital radio, Internet streaming (netcasting/webcasting) and mobile multimedia applications. The limited overall bandwidth available for these systems makes it necessary to use a low bitrate, highly efficient perceptual audio codec in order to create audio that will attract and hold listeners.

In Internet streaming applications, the connection bandwidth that can be established between the streaming server and the listener's client player application depends on the listener's connection to the Internet. In many cases today, people use analog modems or ISDN lines with a limited data rate — lower than the rate that can produce appealing audio quality with conventional perceptual audio codecs. Moreover, even if consumers connect to the Internet through high bandwidth connections such as xDSL, or CATV, the ever-present congestion on the Internet limits the connection bitrate that can be used without audio dropouts and rebuffering. Furthermore, when netcasters pay for bandwidth by the bit, using a highly efficient perceptual codec at low bitrates can make netcasting profitable for the first time.

In mobile communications, the overall bandwidth available for all services in a certain given geographic area (a network cell) is limited, so the system operator must take measures to allow as many users as possible in that network cell to access mobile communication services in parallel. Highly efficient speech and audio codecs allow operators to use their spectrum most efficiently. Considering the impact that the advent of multimedia services has on the data rate demands in mobile communication systems, it is clear that even with 4G LTE, 3GPP, CDMA2000, EDGE, and UMTS, cellular networks will find it necessary to use perceptual codecs at a relatively low data rate. Although many wireless carriers claim to provide high data rates, multimedia requires a consistent data rate to prevent media dropouts. Low bitrate codecs prevent dropouts on congested networks.

Using perceptual codecs at low bitrates, however, has a downside. State-of-the-art perceptual audio codecs such as AAC, achieve "CD-quality" or "transparent" audio quality at a bitrate of approximately 128 kbps (~ 12:1 compression). Below 96 kbps, the perceived audio quality of most of these codecs begins to degrade significantly. Either the codecs start to reduce the audio bandwidth and to modify the stereo image or they introduce annoying coding artifacts caused by a shortage of bits when they attempt to represent the complete audio bandwidth. Both ways of modifying the perceived sound can be considered unacceptable above a certain level. At 64 kbps for instance, AAC either would offer an audio bandwidth of only ~ 12.5 kHz or introduce a fair amount of coding artifacts. Each of these factors severely affects the listening experience and is not perceived as high fidelity.

SBR (Spectral Band Replication) is a very useful audio coding enhancement tool. It can improve the performance of low bitrate audio and speech codecs by either increasing the audio bandwidth at a given bitrate or by improving coding efficiency at a given quality level.

SBR can increase the limited audio bandwidth that a conventional perceptual codec offers at low bitrates so that it equals or exceeds analog FM audio bandwidth (15 kHz). SBR can also improve the performance of narrow-band speech codecs, offering the broadcaster and netcaster speech-only channels with 12 kHz audio bandwidth used, for example, in multilingual broadcasting. As most speech codecs are very band-limited, SBR is important not only for improving speech quality, but also for improving speech intelligibility and speech comprehension. SBR is mainly a post-process, although the encoder performs some pre-processing to guide the decoding process.

From a technical point of view, SBR is a method for highly efficient coding of high frequencies in audio compression algorithms. When used with SBR, the underlying coder is only responsible for transmitting the lower part of the spectrum. The higher frequencies are generated by the SBR decoder, which is mainly a post-process following the conventional waveform decoder. Instead of transmitting the spectrum, SBR reconstructs the higher frequencies in the decoder based on an analysis of the lower frequencies transmitted in the underlying coder. To ensure an accurate reconstruction, some guidance information is transmitted in the encoded bitstream at a very low data rate.

The reconstruction is efficient for harmonic as well as for noise-like components and permits proper shaping in both the time and frequency domains. As a result, SBR allows full bandwidth audio coding at very low data rates and offers significantly increased compression efficiency compared to the core coder.

SBR can enhance the efficiency of perceptual audio codecs by ~ 30% (even more in certain configurations) in the medium to low bitrate range. The exact amount of improvement that SBR can offer also depends on the underlying codec. For instance, combining SBR with AAC achieves a 64 kbps stereo stream whose quality is comparable to AAC at 96 kbps stereo. SBR can be used with mono and stereo as well as with multichannel audio.

SBR offers maximum efficiency in the bitrate range where the underlying codec itself is able to encode audio signals with an acceptable level of coding artifacts at a limited audio bandwidth.

Parametric Stereo

Parametric Stereo is a major technology to enhance the efficiency of audio compression for low bitrate stereo signals. Parametric Stereo is fully standardized in MPEG-4 and is the new component within HE-AACv2. As of today, Parametric Stereo is optimized for the range of 16-40 kbps and provides high audio quality at bitrates as low as 24 kbps.

The Parametric Stereo encoder extracts a parametric representation of the stereo image of an audio signal. Meanwhile, a monophonic representation of the original signal is encoded via AAC+SBR. The stereo image information is represented as a small amount of high quality parametric stereo information and is transmitted along with the monaural signal in the bit stream. The decoder uses the parametric stereo information to regenerate the stereo image. This improves the compression efficiency compared to a similar bit stream without Parametric Stereo.

MPEG Surround

MPEG Surroundstandardized in ISO/IEC 23003-1:2007, is an efficient technology for multi-channel audio compression that extends the concept of “parametric stereo” to more than two channels. Rather than performing a discrete coding of the individual audio input channels, MPEG Surround captures the spatial image of a multi-channel audio signal into a compact set of parameters that are used to synthesize a high quality multi-channel representation from a transmitted downmix signal.

MPEG Surround combines a core audio codec (usually AAC or HE-AAC, although other codecs like MP3 can be used) with a parametric side-channel containing the information necessary to distribute the audio to the various surround output channels. Typically the core codec is stereo (allowing MPEG Surround to be completely stereo-compatible), but a mono core codec can be used at very low bitrates at the expense of reduced subjective performance.

Using Data Compression for Contribution

Using lossy compression to store program material for playout is one area where AM practice might diverge from FM and DAB practice. Because of the lower audio resolution of AM at the typical receiver, an AM station trying to economize on storage might want to use a lower data rate than an FM or DAR station. However, this is likely to be false economy if the owner of this library ever wants to use it for higher fidelity services like netcasting, FM or DAR in the future. In general, increasing the quality reduces the likelihood that the library will cause problems in future.

Any library recorded for general-purpose applications should use at least 44.1 kHz sample rate so that it is compatible with digital radio systems having 20 kHz bandwidth. If the library will only be used on FM and AM, 32 kHz is adequate and will save considerable storage. However, given the rise of digital radio and netcasting, we cannot recommend that any forward-looking station use 32 kHz for storage. Because CD audio is 44.1 kHz, sample rate conversion is unnecessary. This eliminates a process that can potentially degrade the audio quality.

At this writing, the cost of hard disks and other digital storage media is declining so rapidly that there no argument for storing programming using lossy compression, and contribution codecs have thus fallen out of favor. Of course, either no compression or lossless compression will achieve the highest quality. (There is no quality difference between these.). Uncompressed audio workflow is much easier to deal with.

Conversely, cascading stages of lossy compression produces inexorable quality loss and can cause noise and distortion to become unmasked. Multiband audio processing can also cause noise and distortion to become unmasked, because multiband processing “automatically re-equalizes” the program material so that the frequency balance is not the same as the frequency balance seen by the psychoacoustic model in the encoder. Storing audio in linear PCM format makes the audio easier to edit and copy without quality loss.

Many facilities are receiving source material that has been previously processed through a lossy data reduction algorithm, whether from satellite, over landlines, or over the Internet. Sometimes, several encode/decode cycles will be cascaded before the material is finally broadcast. As stated above, all such algorithms operate by increasing the quantization noise in discrete frequency bands. If not psychoacoustically masked by the program material, this noise may be perceived as distortion, “gurgling,” or other interference. Cascading several stages of such processing can raise the added quantization noise above the threshold of masking, such that it is heard.

Accordingly, if you must use lossy data reduction in the studio, you should use the highest data rate possible along with a codec designed for contribution, like MPEG1 Layer2 at 384 kb/sec or Dolby E. This maximizes the headroom between the added noise and the threshold where it will be heard. In addition, you should minimize the number of encode and decode cycles, because each cycle moves the added noise closer to the threshold where the added noise is heard. This is particularly critical if the transmission medium itself (such as DAR, satellite broadcasting, or netcasting) uses lossy compression.

Pitfalls When Using Dolby AC3 as a Contribution Codec

Although the Dolby AC3 codec was designed as a transmission codec, some facilities use it as a contribution codec. (Dolby recommends Dolby E as a contribution codec and deprecates using AC3 for this task.) In addition to the obvious issue of "cascaded codecs," AC3 has a tricky potential pitfall when used as a contribution codec.

AC3 metadata ("data about the data" transmitted as part of the AC3 bitstream) includes "dynamic range control" words. Dolby's intent was to create dynamics compression that is entirely under the listener's control. In essence, the AC3 encoder applies its input signal to a wideband compressor and transmits the compressor's gain control signal along with the uncompressed audio that was applied to the input of the compressor. This way, the listeners can enjoy the full dynamic range of the original signal or can apply compression if they prefer a smaller dynamic range.

To prevent audible wideband gain pumping, the amount of dynamic range compression available from AC3 was purposely limited. Many consumers prefer audio dynamic range to be controlled more tightly so they can enjoy television programs at low volumes and/or in noisy environments. Producing more dynamic range compression without objectionable side effects requires use of multiband compressors like that found in Optimods designed for digital television, such as Optimod-Surround 8685.

When AC3 is used as a contribution codec, it is possible to configure the Dolby AC3 decoder incorrectly so that it applies unwanted and unexpected wideband compression to its output. If your facility is using AC3 as a contribution codec, it is important to double-check the configuration of the decoder to make sure that dynamic range control is not applied to the decoder's output. It is safest to choose a DRC profile of "None" at the AC3 encoder. This ensures that dynamic compression cannot be accidentally applied to the signal in the decoder.

Part 2: System Considerations

Analog Interconnection

For analog connections, we recommend balanced connections between devices using XLR-type connectors for termination because of their robustness. Use twoconductor foil-shielded cable (such as Belden 8451, 1503A, 1504A, 1508A, or equivalent), because signal current flows through the two conductors only. The shield does not carry signal and is used only for shielding. It should be connected at the input only to prevent ground loop hum

Grounding

Very often, grounding is approached in a “hit or miss” manner. Nevertheless, with care it is possible to wire an audio studio so that it provides maximum protection from power faults and is free from ground loops (which induce hum and can cause oscillation).

In an ideal system:

- All units in the system should have balanced inputs. In a modern system with low output impedances and high input impedances, a balanced input will provide common-mode rejection and prevent ground loops—regardless of whether it is driven from a balanced or unbalanced source.
- All equipment circuit grounds must be connected to each other; all equipment chassis grounds must be connected together.
- In a low RF field, cable shields should be connected at one end only—preferably the destination (input) end. This also prevents input noise pick-up when the output is disconnected.
- In a high RF field, audio cable shields should be connected to a solid earth ground at both ends to achieve best shielding against RFI.
- Whenever coaxial cable is used, shields are automatically grounded at both ends through the terminating BNC connectors.

AES3 and SP/DIF Digital Interconnection

Per the AES3 standard, each digital input or output line carries two audio channels. The connection is 110Ω balanced and is transformer-coupled in high-quality equipment.

The AES3 standard specifies a maximum cable length of 100 meters. While almost any balanced, shielded cable will work for relatively short runs (5 meters or less), longer runs require used of 110Ω balanced cable like Belden 1800B, 1801B (plenum rated), multi-pair 180xF, 185xF, or 78xxA. Single-pair Category 5, 5e, and 6 Ethernet cable will also work well if you do not require shielding. (In most cases, the tight balance of Category 5/5e/6 cable makes shielding unnecessary.)

The AES3id standard is best for very long cable runs (up to 1000 meters). This specifies 75Ω unbalanced coaxial cable, terminated in 75Ω BNC connectors. A 110Ω/75Ω balun transformer is required to interface an AES3id connection to an AES3 connection.

S/PDIF[8] is a consumer digital standard closely related to the AES3id standard. However S/PDIF is available in two different physical interfaces, coaxial and optical. Coaxial is is 75Ω unbalanced and optical is TOSLINK. Both interfaces offer excellent quality and are good for short distances. Format converters are available to go between either coaxial or optical and/or AES3.

Digital Audio Clock

Digital audio requires an audio clock, which allows downstream devices to reconstruct the sample frequency. The audio clock is also referred to as "sample clock." It is a simple frequency reference and carries no time-of-day information; it is not the same as timecode. It is typically distributed as "wordclock" (typically a 5V p-p squarewave at the system sample frequency) or AES 11.

Published by the Audio Engineering Society, the AES11 standard provides a systematic approach to the synchronization of digital audio signals[9]. Recommendations are made concerning the accuracy of sample clocks as embodied in the interface signal and the use of this format as a convenient synchronization reference where signals must be rendered co-timed for digital processing. Synchronism is defined, and limits are given which take account of relevant timing uncertainties encountered in an audio studio.

AES11 recommends using an AES3 signal to distribute audio clocks within a facility. In this application, the connection is referred to as a Digital Audio Reference Signal (DARS).

AES11 Annex D (in the November 2005 or later printing or version) shows an example method to provide isochronous timing relationships for distributed AES3 structures over asynchronous networks such as AES47 where reference signals may be locked to common timing sources such as GPS.

In addition, the Audio Engineering Society has now published a related standard called AES53 that specifies how the timing markers already specified in AES47 may be used to associate an absolute time-stamp with individual audio samples. This may be closely associated with AES11 and used to provide a way of aligning streams from disparate sources, including synchronizing audio to video in networked structures.

The media profile defined in annex A of AES67 provides a means of using AES11 synchronization via the Precision Time Protocol (PTP),[10] which is a protocol used to

---

[8] S/PDIF is standardized in IEC 60958 as IEC 60958 type II (IEC 958 before 1998)

[9] http://www.aes.org/publications/standards/search.cfm?docID=18

[10] The above text regarding AES11, AES47, AES53, and AES67 was retrieved from https://en.wikipedia.org/wiki/AES11 24 October 2018.

synchronize clocks throughout an Ethernet network. On a local area network, it achieves clock accuracy in the sub-microsecond range. In 2002 PTP originated with the IEEE 1588-2002 standard, which was later updated to IEEE 1588-2008. This defines PTP v2 with improved precision and robustness; however, it is not backward-compatible with the 2002 version.

By default, digital audio devices typically use audio inputs such as AES3 or S/PDIF to recover the clock from the digital audio bitstream. However, in larger installations where multiple digital audio inputs are used, it usually desirable to use a master clock system to lock all digital audio outputs to the same digital audio clock reference. This is useful for several reasons:

- REDUNDANCY – If the digital audio input that is used for clock reference is lost, then all of the remaining inputs have also lost their digital audio clock reference. Using a separate digital audio clock reference and the appropriate digital audio clock input, such as wordclock or AES3 DARS eliminates such problems.
- RELIABILITY – Without a separate clock reference signal, the last digital audio device in a chain can only receive an accurate clock if every intermediate audio device extracts the clock accurately and passes it correctly to the next device. A separate clock reference ensures that each device receives the reference independently.
- PRECISE SAMPLE-RATE – Using one high-accuracy, preferably GPS-based, digital clock reference for all streams ensures that sample frequencies in all part of the system are identical. Using the same timebase for the encoder timestamps also guarantees synchronicity between presentation time and audio sample rate to eliminate drift at the client player, assuming the player client has been written to take advantage of this. This plays audio at the correct speed.
- LOWER JITTER – While the all-digital parts of a system are unaffected by clock jitter unless it is so bad that it is impossible to accurate extract the digital data, clock jitter is crucially important at the analog interfaces: the A/D and D/A converters. Clock jitter at the converters causes phase modulation of the audio, which adds unpleasant-sounding inharmonic sidebands to the audio spectrum. Jitter-induced sidebands in the A/D converter irrevocably contaminate the audio; such distortion cannot be removed by downstream processing. Conversely, at the D/A one can use jitter reduction techniques (like a narrowband phase-locked loop) to reduce the jitter in the received clock before it is applied to the converter.
- A clock extracted from an AES3 or S/PDIF audio signal is inherently vulnerable to jitter due to the signal's audio modulation. Although this can be reduced by post-processing the extracted clock, jitter reduction requires addi-

tional circuitry that increases cost. Independent clock reference signals have no inherent jitter because they are not mixed with audio modulation; the jitter is limited only by the quality of hardware in the master clock generator.

- BIT-ACCURACY - Although not necessarily important for broadcast or streaming applications, recording and production facilities that use linear PCM and don't use coded audio may rely on a precision digital audio clock reference so that all digital equipment is synchronous and bit-accurate. This avoids asynchronous sample rate conversion, which can introduce subtle artifacts.

Sample Rate Conversion in Digital Audio Distribution

Digital systems that do not have sample rate converters in their signal path are considered synchronous if there is no further digital signal processing. As described above, a master clock signal, typically in wordclock or AES11 format, is required for all devices in the audio signal path to remain in sync. To be bit-accurate, a system must be synchronous and must not change the word length of the audio data being conveyed.

Digital systems that have sample rate converters in their signal path are not bit-accurate and are considered asynchronous even though they can also be locked to a master clock. The advantage to asynchronous digital audio is that just about anything can be connected to just about anything else without difficulty. However, there are cautions.

Computer audio can be very confusing, especially when one is trying to achieve bit-accurate audio. Because audio has many different sources and destinations, and because normal computer motherboards provide no way to lock a computer's internal timebase to an external reference, computer operating systems contain software sample rate converters for both record and play to make things easier for the average user. For the professional, this can cause performance problems and sample rate conversion can occur when it is not wanted or expected.

To work around this problem, most professional users use the ASIO interface and drivers. However, not all audio applications support ASIO. Furthermore, if recording or encoding from another application is required, this is not supported by ASIO. In principle, it is possible to use Windows' built-in WASAPI (Windows Audio Session API)[11] Exclusive Mode[12] as an alternative to ASIO, but the applications you are using must support this.

To prevent sample rate conversion from occurring, be certain that source sample rates match record sample rates and that play sample rates match destination sam-

---

[11] https://docs.microsoft.com/en-us/windows/desktop/coreaudio/wasapi

[12] https://docs.microsoft.com/en-us/windows/desktop/coreaudio/exclusive-mode-streams

ple rates. Microsoft Windows has a checkered history regarding its built-in sample rate conversion. Microsoft Windows XP had a high quality sample rate converter in the audio stack for both record and play, and if sample rate conversion was unavoidable, Windows XP provided a high-quality conversion.

Microsoft implemented a new audio stack in Windows Vista/7. Legacy audio applications use the emulation mode of this audio stack. In early versions of Windows 7, the record sample rate converter had poor performance, while the play sample rate conversion appears fine.

In Windows 8.1, Microsoft fixed the record sample rate converter, and following pressure from Orban and others, also made the fix available for Windows 7 as a hotfix: http://support.microsoft.com/kb/2653312

In Windows 10, the performance of the record and playback sample rate converters is satisfactory.

Audio over IP

Audio-Over-IP (AoIP) digital audio connections are becoming progressively more popular because they simplify wiring and allow use of off-the-shelf Ethernet switches, reducing implementation cost. Popular variants include Axia Livewire, RAVENNA, Wheatstone WheatNet, and Audinate Dante. Each manufacturer specifies recommended networking hardware that has proven reliable with its system and it is wise to follower a given manufacturer's advice; failure to do so can cause glitches and dropouts.

Audio over Ethernet is available as Layer 2 and Layer 3 protocols. Layer 2 is not TCP/UDP and cannot be routed with common IP routers. The most common Layer 2 protocol is ABV, whose main advantage is very low latency.

The Audio Engineering Society has published to AES67 standard, which describes a standardized audio transport mechanism that allows basic audio interconnection between AOiP networks. AES67 specifies interoperability requirements for Layer 3 connections. These connections use UDP RTP multicast for audio packets and TCP RTSP for stream initialization. AES67 uses PTP v2 clock for synchronization, so not all grandmaster clocks will work.

AES67 also includes requirements for interoperability of AVB (Layer 2) networks, which must be routed using specialized routers and switches.

AES67 does not include "discovery," which is the ability of a given network to automatically discover devices connected it and to configure itself appropriately, or control protocols that allow devices on the network to be remote-controlled. AES70 ("AES standard for audio applications of networks — Open Control Architecture") attempts to fill in this gap by defining a scalable control-protocol architecture for professional media networks. AES70 addresses device control and monitoring only; it does not define standards for streaming media transport. However, the Open Control Architecture (OCA) is intended to cooperate with various media transport architectures.

At this writing, the various proprietary AoIP systems are moving towards AES67 and AES70 compatibility, and the curious reader should visit the manufacturers' websites to ascertain their current progress toward interoperability using these standards, as well as other details of these complex products.

There are several other proprietary protocols. We expect these to fade away over time because of AES67's potential for achieving universal interoperability between various manufacturers' products. At this writing RAVENNA and Axia/Livewire V2 use AES67 as their primary protocol. Dante and Wheatnet support it, but not as a primary protocol.

Networked audio connections follow the same exact wiring convention as all other Ethernet data networks. Category 5, 5e, or 6 Ethernet cable should be used.

Regarding the audio quality of AOiP connections, there is little to say. Because the connections are digital, the main dangers are clicks, pops, and audio dropouts caused by data loss in transport. There are many potential sources of such difficulties, including network overloads, networks switches that are incompatible with your chosen AoIP system, and mismatched bitclocks between sources and destinations that can cause buffer underflows or overflows. A second potential source of quality degradation is poor-quality sample rate conversion, although this is rare with today's high-performance integrated SRC chips.

MADI

Multichannel Audio Digital Interface (MADI) or AES10 is an Audio Engineering Society standard that defines the data format and electrical characteristics of a point-to-point interface that carries multiple channels of digital audio. The AES first documented the MADI standard in AES10-1991, and updated it in AES10-2003 and AES10-2008. The MADI standard includes a bit-level description and has features in common with the two-channel AES3 interface.

MADI supports serial digital transmission over coaxial cable or fiber-optic lines of 28, 56, or 32, 64 channels; and sampling rates of up to 192 kHz with an audio bit depth of up to 24 bits per channel. Like AES3 and ADAT Lightpipe it is a unidirectional interface from one sender to one receiver.[13]

AES50 is a further development of the AES10 protocol. It enables very low-latency point-to-point transport up to 100 meters using CAT5 Ethernet cables and Ethernet

---

[13] https://en.wikipedia.org/wiki/MADI

switchers (but *not* routers). It is based on the OSI-Layer 1 protocol, so it does not use the Ethernet frame structure. The standard supports separate transmission of wordclock and full-duplex transmission up of to 48 channels. Because it is not routable, it does not qualify as an audio networking protocol like AES67.

Headroom and Metering

One of the most misunderstood details of audio is exactly how to measure levels, and how analog levels relate to digital levels. Figure 1 on page 30 shows the calibration and level relationships between the following meters, where all meters are displaying a sinewave at SMPTE RP155 reference level of –20 dBFS:

- True Peak-Reading Digital Level Meter
- VU Meter (ANSI)
- Peak Program Meter (EBU IEC 268-10 IIB)
- Peak Program Meter (UK IEC 268-10 IIA)
- Peak Program Meter (Nordic IEC 268-10 I)
- Peak Program Meter (DIN IEC 268-10 DIN 45406)

Peak Normalization in Audio Editing Programs

Many audio editing programs permit a sound file to be "normalized," which amplifies or attenuates the level of the file to force the highest peak to reach 0 dBFS. This is unwise for several reasons. Peak levels have little to do with loudness, so normalized files are likely to have widely varying loudness levels depending on the typical peak-to-average ratio of the audio in the file. Also, if any processing occurs after the normalization process (such as equalization), one needs to ensure such processing does not clip the signal path. If the processing adds level, one must compensate by applying attenuation before the processing to avoid exceeding 0 dBFS, or by using floating point arithmetic. If attenuation is applied, one must use care to ensure that

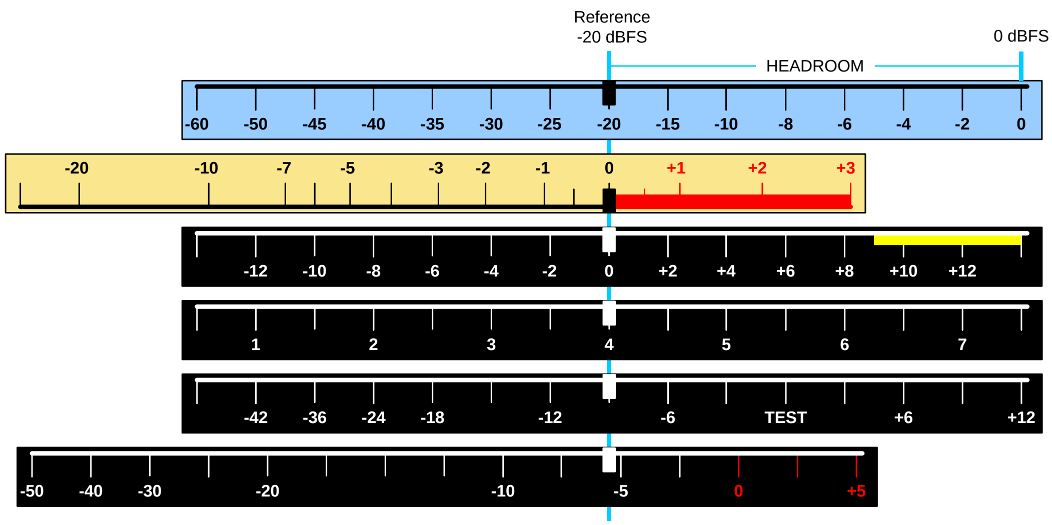

Figure 1: Comparison of Various Level Meter Scales

the attenuated signal remains adequately dithered (see page 44).

Moreover, normalization algorithms often do not use true peak level as specified in ITU Recommendation BS.1770, or oversampling. If they do not, files normalized by the algorithms can clip downstream D-A and sample rate converters due to the 0dBFS+ phenomenon (see page 64), and cause more distortion and aliasing.

The audio processor analog input A/D should clip at the same audio level as the source amplifier or console. The input level should be adjusted so this clip level is the same.

Subjective loudness metering is discussed in *Measuring and Controlling Loudness* starting on page 35.

Headroom

The single most common cause of distorted broadcast/netcast sound is probably clipping—intentional (in the audio processing chain) or unintentional (in the program chain). In order to achieve the maximum benefit from processing, there must be *no* clipping before the processor! The gain and overload point of *every* electronic component in the station must therefore be critically reviewed to make sure they are not causing clipping distortion or excessive noise.

In media with limited dynamic range (like magnetic tape), small amounts of peak clipping introduced to achieve optimal signal-to-noise ratio are acceptable. Nevertheless, there is no excuse for *any clipping at all* in the purely electronic part of the signal path, since good design readily achieves low noise and wide dynamic range.

Check the following components of a typical audio facility for operating level and headroom:

- Analog-to-digital converters
- Studio-to-transmitter link (land-line, microwave, or optical fiber)
- Microphone preamps
- Console summing amplifiers
- Line amplifiers in consoles, tape recorders, etc.
- Distribution amplifiers (if used)
- Signal processing devices (such as equalizers)
- Specialized communications devices (including remote broadcast links and telephone interface devices)
- Phono preamps
- Tape and cart preamps

- Record amplifiers in tape machines
- Computer sound cards

VU meters are worthless for checking peak levels. Even peak program meters (PPMs) are insufficiently fast to indicate clipping of momentary peaks because their integration time is 5 or 10ms, depending on which variant of the PPM standard is employed. While PPMs are excellent for monitoring operating levels where small amounts of peak clipping are acceptable, the peak signal path levels should be monitored with a *true* peak-reading meter or oscilloscope. Particularly, if they are monitoring pre-emphasized signals, PPMs can under-read the true peak levels by 5dB or more. Adjust gains so that peak clipping *never* occurs under any reasonable operating conditions (including sloppy gain riding by the operator).

It is important to understand that digital "peak-reading" meters, also known as "bit meters", may show the peak value of digital samples in a bitstream without correctly predicting the peak level of the reconstructed analog waveform after D/A conversion or the peak level of digital samples whose sample rate has been converted. The meter may under-read the true peak level by as much as 3 dB. This phenomenon is known as 0dBFS+. The ITU BS.1770 Recommendation ("Algorithms to measure audio programme loudness and true-peak audio level") suggests oversampling a true peak reading meter by at least 4x and preferably 8x. By filling in the "space between the samples," oversampling allows the meter to indicate true peaks more accurately. This allows the 0 dBFS+ phenomenon to be monitored and prevented. See *0 dBFS+* on page 64.

For older equipment with very soft clipping characteristics, it may be impossible to see a well-defined clipping point on a scope. Or, worse, audible distortion may occur many dB below the apparent clip point. In such a case, the best thing to do is to determine the peak level that produces 1% THD, and to arbitrarily call *that* level the clipping level. Calibrate the scope to this 1% THD point, and then make headroom measurements.

Engineers should also be aware that certain system components (like microphone preamps, phono preamps, and computer soundcards) have absolute *input* overload points. Difficulties often arise when gain controls are placed *after* early active stages, because the input stages can be overloaded without clipping the output. Many broadcast microphone preamps are notorious for low input overload points, and can be easily clipped by high-output microphones and/or screaming announcers. Similar problems can occur inside consoles if the console designer has poorly chosen gain structures and operating points, or if the "master" gain controls are operated with unusually large amounts of attenuation.

When operating with nominal line levels of +4 or +8dBu, the *absolute* clipping point of the line amplifier becomes critical, The headroom between nominal line level and the amplifier clipping point should be greater than 16dB. A line amplifier for a +4dBu line should, therefore, clip at +20dBu or above, and an amplifier for a +8dBu line should clip at +24dBu or above. IC-based equipment (which almost always clips at +20dBu or so unless transformer-coupled) is not suitable for use with +8dBu lines. +4dBu lines have become standard in the recording industry, and are preferred for all new studio construction (recording or broadcast) because of their compatibility with IC opamp operating levels.

Some consumer and "semi-pro" equipment uses a nominal line level of –10 dBV. To prevent clipping, *this equipment can only be used in +4 dBu environments if attenuation is applied prior to the input*. Sometimes, an analog-domain volume control in the equipment provides this, but this does not help if there are any active (amplification) devices before the volume control. If the equipment digitizes the input signal, note carefully that a *level control in the digital domain cannot eliminate clipping that occurs earlier in the analog domain or the A/D converter*.

The same headroom considerations that apply to analog also apply to many digital systems. The only digital systems that are essentially immune to such problems are those that use floating point numbers to compute and distribute the digital data. While floating point arithmetic is relatively common within digital signal processors, mixers, and digital audio workstations, it is very uncommon in external distribution systems. The core audio input/output of a computer's operating system is usually fixed-point.

Even systems using floating-point representation are vulnerable to overload at the A/D converter. If digital recording is used in the facility, bear in mind that the overload point of digital audio recorders (unlike that of their analog counterparts) is abrupt and unforgiving. *Never* let a digital recording go "into the red"—this will almost assuredly add audible clipping distortion to the recording. Similarly, digital distribution using the usual AES3 connections has a very well defined clipping point—digital full-scale (0 dBFS)—and attempting to exceed this level will result in distortion that is even worse-sounding than analog clipping because the clipping harmonics above one-half the sampling frequency will fold around this frequency, appearing as aliasing products.

0 dBFS is *not* at all the same as 0 VU or 0 PPM! (See *Figure 1: Comparison of Various Level Meter Scales* on page 30.) In a *contribution* system with adequate headroom, 0 VU ("reference level") should be placed at –20 dBFS (SMPTE RP155 standard) or –18 dBFS (EBU R68 standard)[14]. In a *transmission* system where the audio will be transmitted via the Dolby AC3 codec, 0 VU is often placed even lower (typically –24 or –25 dBFS) and the value of Dolby AC3 Dialnorm metadata transmitted to consumers is set to match this reference level. The consumer's receiver then uses the received value of Dialnorm to adjust a "hidden volume control" in series with the volume control available to the consumer. Use of Dialnorm thus allows the loudness of programs from various providers and sources to be consistent regardless of their choice of reference level.

[14] ITU-R BS.1726 ("Signal level of digital audio accompanying television in international programme exchange") allows for use of either the EBU or SMPTE reference levels.

Many systems use digital audio sound cards to get audio signals in and out of computers that are used to store, process, and play audio. However, not all sound cards have equal performance, even when using digital input and output. For example, a sound card may unexpectedly change the level applied to it. Not only can this destroy system level calibration, but gain can introduce clipping and loss can introduce truncation distortion unless the gain-scaled signal is correctly dithered. If the analog input is used, gain can also introduce clipping, and in this case, loss can compromise the signal-to-noise ratio. Further, the A/D conversion can introduce nonlinear distortion and frequency response errors.

In almost all modern professional facilities, analog reference level = +4 dBu (1.228 V RMS) and circuits clip at +20 dBu or higher. When using analog I/O, consumer and prosumer computer sound cards require input attenuation and output amplification to interface to professional levels. Do not use the software volume control to control input levels; this cannot prevent the input A/D converter from clipping. Align the software output level control by setting the control as high as possible without clipping when a 0 dBFS tone file is played.

Many computer sound card software drivers are incompetently written and do not handle audio levels correctly. To achieve professional results, choose computer sound devices very carefully.

Several sound cards and USB audio devices have a reversed left and right audio clock that causes bit-slip. The resulting digital audio is not correctly time-aligned, which causes an interchannel phase shift that increases with frequency. The sum of the left and right channels does not exhibit a flat frequency response.

The amount of attenuation at a given frequency depends upon the sample rate. A one-sample slip at 32kHz sample rate produces a notch at 16kHz and almost 6dB of loss at 10kHz; 44.1kHz produces almost –3dB at 10kHz and –6dB at 15kHz; 48kHz produces –2dB at 10kHz and –5dB at 15kHz. Because one-sample slip is audible in the mono sum, devices with this problem are inappropriate for broadcast audio applications, especially for mastering a library. Many of these devices were based upon a Texas Instruments USB Codec chip that had its hardware clock reversed. TI has acknowledged the problem and has released revised parts. However, many audio interfaces and codecs in use have this problem and should be scrupulously avoided.

Level metering in sound cards is highly variable. Average, quasi-peak, and peak responses are all common and often inadequately or incorrectly documented (see *Headroom and Metering* on page 30). This is relevant to the question of line-up level. EBU R68 specifies reference level as –18dBfs, while SMPTE RP 155 specifies it as –20dBfs. Unless the sound card's metering is accurate, it is impossible to ensure compliance with the standards maintained within your facility. Many professional sound cards have adequate metering, while this is far less common on consumer sound cards. Further, consumer sound cards often cannot accommodate professional analog levels, balanced lines, or AES/EBU inputs and outputs.

Another potential problem occurs in metering if the signal path prior to a true-peak-reading meter is wholly or partially fixed-point such that peaks above 0 dBFS are clipped off. For example, this can occur if a player codec uses fixed-point arithmetic. In this case, a meter at the codec's output will not be able to correctly read codec-induced peak overshoots above 0 dBFS.

Measuring and Controlling Loudness

Loudness is subjective: it is the intensity of sound as perceived by the ear/brain system. No simple meter, whether peak program meter (PPM) or VU, provides a reading that correlates well to perceived loudness. A meter that purports to measure loudness must agree with a panel of human listeners.

BS.1770 Loudness Meter

In 2006, the ITU-R published Recommendation ITU-R BS.1770: "Algorithms to measure audio programme loudness and true-peak audio level." Developed by G.A. Soulodre, the original BS.1770 loudness meter uses a frequency-weighted RMS measurement intended to be integrated over several seconds — perhaps as long as an entire program segment. As such, it is considered a "long-term" loudness measurement because it does not take into account the loudness integration time constants of human hearing, as does the CBS meter.

A major disadvantage of the BS.1770-1 meter is that it weights silence and low-loudness material the same as high loudness material. This will cause the meter to under-read program material (like dialog) having substantial pauses that contain only low-level ambience because louder program material contributes most to a listener's perception of overall program loudness.

To address this problem, the BS.1770-2 algorithm adds gating to the BS.1770-1 algorithm so that the meter ignores silence and is weighted toward louder program material, which contributes most to a listener's perception of loudness. BS.1770-2 (and higher) indicates only sounds that fall within a floating window that extends from the loudest sounds within the preset integration period to sounds that are 10 dB quieter than the loudest sounds. There are two steps in the gating process: first, an absolute gate removes silent passages; second, a relative gate weights louder parts of the program more heavily that quieter parts.

A more detailed explanation of the algorithm is this:

1. Using the BS.1770-1 algorithm, (i.e., a K-weighting filter followed by RMS summation and averaging), calculate the RMS value in a 400 ms time window. One number is computed for every 400 ms time window. Start computing a new 400ms window every 100 ms, so there is 75% time overlap between windows. Continue computing the RMS values of new 400ms windows throughout the entire duration of the measurement and store all of these results — one number for each 400ms window.

2. If any 400ms window has a value below –70 LKFS, throw it away.

3. Compute the average of the remaining windows over the total time period of the measurement. If any window is less than 10 dB below this average, throw it away.

4. Compute the average of the remaining windows. Display this reading on the meter.

In 2015, the BS.1770 standard was updated to BS.1770-4, which extends the standard to measure the loudness of 7.1 surround and "with-height" systems like Dolby Atmos and MPEG-H.

EBU R128

In August 2010, the EBU published its Loudness Recommendation EBU R128. It specifies how broadcasters and netcasters can measure and normalize audio using Loudness meters instead of Peak Meters (PPMs) or VU meters only, as has been common practice.

EBU R128 is the result of two years of intense work by the audio experts in the EBU PLOUD Group. The new Recommendation is accompanied by a Loudness Metering specification (EBU Tech 3341), a Loudness Range descriptor (EBU Tech 3342), Loudness test material (various different sequences) Production Guidelines (EBU Tech 3343) and Distribution Guidelines (EBU Tech 3344). An EBU Technical Review Article describing the fundamental change in audio in broadcasting is also available: *On the way to Loudness Nirvana*.

EBU R128 recommends normalizing audio at −23 LUFS ±0.5 LU (±1 LU for live programs), measured using the BS.1770-2 (gated) algorithm or higher. The metering approach can be used with virtually all material.

To make sure meters from different manufacturers provide the same reading, EBU Tech 3341 specifies the 'EBU Mode', which includes a *Momentary* (400 ms), *Short Term* (3s) and *Integrated* (from start to stop) meter. The Momentary and Short-Term meters do not use gating; this is only used in the Integrated measurement.

In our opinion, a target loudness −23 LUFS is too low for certain applications. For example, as of 2018 streams normalized to −23 LUFS cannot produce satisfying listening levels on Apple iPhones because the range of the iPhone's volume control is insufficient. We expect that consumer electronics manufacturers will correct this problem in the future, but a decision to set a particular target loudness should also take into account the presence of legacy devices. Following the AES TD1004.1.15-10 recommendation (see *Loudness Balance between Speech and Music* on page 38), we believe that a more realistic target loudness for current player devices like iPhones is −16 LUFS. This allows very high subjective quality while also allowing the program to be played at a level that satisfies listeners.

ATSC A/85

In 2009, the Advanced Television Systems Committee (ATSC) in the United States released a Recommended Practice: *Techniques for Establishing and Maintaining Audio Loudness for Digital Television* (A/85:2009). This was later updated as A/85:2013. A/85 specifies use of the latest version of the Integrated ITU BS.1770 algorithm for measuring the loudness of DTV broadcasts.

In December 2011, the FCC[15] adopted rules implementing the CALM Act[16], which, by law, forbids commercials from being louder than non-commercial program material. The new FCC rules incorporated ATSC A/85 (and, by implication, the BS.1770 meter) as an objective means of verifying that the rule was being obeyed.

The most important difference between R128 and A/85 is that R128 recommends that the target loudness should be measured across all program material, while A/85 recommends that it be measured on the "anchor element," which is usually dialog. In addition, R128 suggests a target loudness of −23 LUFS, while A/85 suggests a target loudness of −24 LUFS.

The ATSC A/85, ITU-R BS.1770, and EBU R128 documents are available as free downloads and their current versions can easily be located with a search engine.

Orban Loudness Meter

Orban now offers a loudness meter application for Windows XP and higher, and for Mac. It is available for free from www.orban.com/meter.[17]

The Orban Loudness Meter receives a stereo or surround (up to 7.1 channels) signal from any Windows sound device and measures its loudness and level. It can simultaneously display instantaneous true peaks (as seen after a D/A converter), digital sample peaks, VU, PPM, CBS Technology Center loudness, ITU BS.1770 loudness, and EBU R 128 Loudness Range. The meter includes peak-hold functionality that makes the peak indications of the meters easy to see. The software has the ability to analyze audio and the audio parts of video files offline for their BS.1770-4 Integrated Loudness, EBU R 128 LRA, highest reconstructed peak level, and number of reconstructed peaks above 0 dBFS. It will graph the BS.1770-4 Integrated Loudness and peak swings of the CBS Loudness Meter as a function of time, and can display a histogram of the BS.1770-4 Integrated Loudness.

---

[15] Federal Communications Commission: the U.S. broadcast regulation agency

[16] The CALM Act applies only to U.S. broadcasters. cable providers, and satellite providers.

[17] Refer to the USER MANUAL HERE link at www.orban.com/meter for up-to-date documentation for the meter.

Jones & Torick (CBS Technology Center) Meter

The CBS meter is a "short-term" loudness meter intended to display the details of moment-to-moment loudness with dynamics similar to a VU meter. It uses the Jones & Torick algorithm[18]. Orban's DSP implementation of this algorithm (used in the free Orban Loudness Meter software and in several Optimod products, including Optimod-PCn 1600 software) typically matches the original meter within 0.5 dB on sinewaves, tone bursts and noise. (The original meter uses analog circuitry and an LED bar graph display with 0.5 dB resolution.) Many researchers have been curious about the Jones & Torick meter but been unable to evaluate it and compare it with other loudness meters. Orban developed this software because we believed it would be useful to practicing sound engineers and researchers.

The Jones & Torick algorithm improves upon the original loudness measurement algorithm developed by CBS researchers in the late 1960s. Its foundation is psychoacoustic studies done at CBS Laboratories over a two year period by Torick and the late Benjamin Bauer. After surveying existing equal-loudness contour curves and finding them inapplicable to measuring the loudness of broadcasts, Torick and Bauer organized listening tests that resulted in a new set of equal-loudness curves based on octave-wide noise reproduced by calibrated loudspeakers in a semireverberant 16 x 14 x 8 room, which is representative of a room in which broadcasts are normally heard. They published this work in "Researches in Loudness Measurement," IEEE Transactions on Audio and Electroacoustics, Volume AU-14, Number 3, September 1966, pp. 141-151. This paper also presented results from other tests whose goal was to model the loudness integration time constants of human hearing.

Orban has written a white paper comparing the CBS and BS.1770 meters:
http://www.orban.com/support/orban/techtopics/White_Paper-BS_1770_vs_CBS_meter.pdf

Loudness Balance between Speech and Music

The VU meter is very deceptive when indicating the balance between speech and music. The most artistically pleasing balance between speech and music is usually achieved when speech is peaked 4–6dB lower than music on the console VU meter. If heavy processing is used, the difference between the speech and music levels may have to be increased. Following this practice will also help reduce the possibility of clipping speech, which is much more sensitive to clipping distortion than is most music.

If a PPM is used, speech and music should be peaked at roughly the same level. However, please note that what constitutes a correct "artistic balance" is highly subjective, and different listeners may disagree strongly. Each broadcasting organiza-

---

[18] Jones, Bronwyn L.; Torick, Emil L., "A New Loudness Indicator for Use in Broadcasting," J. SMPTE September 1981, pp. 772-777.

tion has its own guidelines for operational practice in this area. So the suggestions above are exactly that: just suggestions.

For a given VU or PPM indication, the loudness of different talkers and different music may vary significantly. A short-term loudness meter like the Jones & Torick meter can help operators maintain appropriate voice/music balance by estimating more accurately than a PPM or VU the actual loudness of each program segment.

The BS.1770 Integrated loudness meter can cause inartistic speech/music balances if speech and music are normalized to the same target loudness (see *BS.1770 Loudness Meter* on page 35). In 2015 the Audio Engineering Society released AES TD1004.1.15-10: *Recommendation for Loudness of Audio Streaming and Network File Playback.* This states:

> Within a given program, the largest perceived difference to be noted is speech versus music. Speech normalized to the same Integrated Loudness as a music stream inevitably sounds too loud. It is recommended to normalize speech (dialog) segments within other segments 2 to 4 LU (or more) below the loudness of the other segments.

Many of Orban's Optimod audio processors have automatic speech/music detection and can automatically change processing parameters for speech and music. Setting these parameters to achieve your organization's desired speech/music balance provides an effective way of controlling this balance automatically.

Replay Gain

A popular means of estimating and controlling the loudness of audio files is the Replay Gain[19] technique. The computes a gain factor to be applied to the file when played back; this gain factor is stored as metadata in the file header. The goal is to achieve consistent long-term loudness from track to track. The gain factor is computed by the following steps:

1. Equal Loudness Filtering

The human ear does not perceive sounds of all frequencies as having equal loudness. For example, a full scale sine wave at 1kHz sounds much louder than a full scale sine wave at 10kHz, even though the two have identical energy. To account for this, the signal is filtered by an inverted approximation to the equal loudness curves (sometimes referred to as Fletcher-Munson curves).

2. RMS Energy Calculation

Next, the energy during each moment of the signal is determined by calculating the Root Mean Square of the waveform every 50ms.

---

[19] http://replaygain.hydrogenaudio.org/index.html

3. Statistical Processing

Where the average energy level of a signal varies with time, the louder moments contribute most to our perception of overall loudness. For example, in human speech, over half the time is silence, but this does not affect the perceived loudness of the talker at all! For this reason, the RMS values are sorted into numerical order, and the value 5% down the list is chosen to represent the overall perceived loudness of the signal.

4. Calibration with reference level

A suitable average replay level is 83dB SPL. A calibration relating the energy of a digital signal to the real world replay level has been defined by the SMPTE. Using this calibration, we subtract the current signal from the desired (calibrated) level to give the difference. We store this difference in the audio file.

5. Replay Gain

The calibration level of 83dB can be added to the difference from the previous calculation, to yield the actual Replay Gain. NOTE: we store the differential, NOT the actual Replay Gain.

Electronic Quality

Assuming that the transmission does not use excessive lossy compression, digital audio broadcasting and netcasting have the potential for transmitting the highest subjective quality to the consumer and require the most care in maintaining audio quality in the transmission plant. This is because these transmission media do not use pre-emphasis and have a high signal-to-noise ratio that is essentially unaffected by reception conditions. The benefits of an all-digital facility using minimal (or no) lossy compression prior to transmission will be most appreciated in DAB/HD/SAT Radio and netcasting services.

FM has four fundamental limitations that prevent it from ever becoming a transmission medium that is unconditionally satisfying to “golden-eared” audiophiles. These limitations must be considered when discussing the quality requirements for FM electronics. The problems in analog disc and tape reproduction discussed in the Appendix to this document are much more severe by comparison, and the subtle masking of basic FM transmission limitations is irrelevant to those discussions. AM quality at the typical receiver is far worse, and “golden ear” considerations are completely irrelevant because they will be masked by the limitations of the receivers and by atmospheric and man-made noise.

The four FM quality limitations are these:

A) **Multipath distortion**. In most locations, a certain amount of multipath is unavoidable, and this is exacerbated by the inability of many apartment-dwellers to use rotor-mounted directional antennas.

B) The FM stereo multiplex system has a "**sample rate**" of 38 kHz, so its bandwidth is theoretically limited to 19 kHz, and practically limited by the characteristics of "real-world" filters to between 15 and 17 kHz.

C) **Limited IF bandwidth** is necessary in receivers to eliminate adjacent and alternate channel interference. Its effect can be clearly heard by using a tuner with switch-selectable IF bandwidth. Most stations cannot be received in "wide" mode because of interference. But if the station is reasonably clean (well within the practical limitations of current broadcast practice) and free from multipath, then a clearly audible reduction in high-frequency "grit" or distortion is heard when switching from "normal" to "wide" mode.

D) Depending on the Region, FM uses either 50μs or 75μs **pre-emphasis**. This severely limits the power-handling capability and headroom at high frequencies and requires very artful transmission processing to achieve a bright sound typical of modern CDs. Even the best audio processors compromise the quality of the high frequencies by comparison to the quality of "flat" media like DAB, HD and satellite radio.

These limitations have considerable significance in determining the cost effectiveness of current broadcast design practice.

Most older broadcast electronic equipment (whether tube or transistor) is measurably and audibly inferior to modern equipment. This is primarily due to a design philosophy that stressed ruggedness and RFI immunity over distortion and noise, and to the excessive use of poor transformers. Frequency response was purposely rolled off at the extremes of the audio range to make the equipment more resistant to RFI. Cascading such equipment tends to increase both distortion and audible frequency response rolloffs to unacceptable levels.

Modern analog design practice emphasizes the use of high slew rate, low-noise, low-cost IC operational amplifiers such as the National LM4562 family, the Signetics NE5534 family, the National LF351 family and the Texas Instruments TL070 family. When the highest quality is required, designers will choose premium-priced opamps from Analog Devices, Linear Technology and Burr Brown, or will use discrete class-A amplifiers. However, the 5532 and 5534 can provide *excellent* performance when used properly, and it is hard to justify the use of more expensive amplifiers except in specialized applications like microphone preamps, active filters, and composite line drivers. While some designers insist that only discrete designs can provide ultimate quality, the performance of the best of current ICs is so good that discrete designs are just not cost effective for broadcast/netcast applications—especially when the basic FM, DAB, audio codec quality limitations are considered.

Some have claimed that **capacitors** have a subtle, but discernible effect upon sonic quality. Polar capacitors such as tantalums and aluminum electrolytics behave very differently from ideal capacitors. In particular, their very high dissipation factor and dielectric absorption can cause significant deterioration of complex musical waveforms. Ceramic capacitors have problems of similar severity. Polyester film capacitors can cause a similar, although less severe, effect when audio is passed through them. Accordingly, DC-coupling between stages is best (and easy with opamps operated from dual-positive and negative power supplies). Coupling capacitors should be used only when necessary (for example, to keep DC offsets out of faders to prevent "scratchiness"). If capacitors must be used, polystyrene, polypropylene, or polycarbonate film capacitors are preferred. If electrolytic capacitors are used, it is wise to use them with DC bias so that AC audio signals can never reverse-bias them.

To eliminate DC offsets, the best audio designs use servos instead of coupling capacitors. However, if it is impractical to eliminate capacitors or to change capacitor types, do not be too concerned: it is probable that other quality-limiting factors will mask the capacitor-induced degradations.

Of course, the number of **transformers** in the audio path should be kept to an absolute minimum. However, transformers are sometimes the only practical way to break ground loops and/or eliminate RFI. If a transformer is necessary, use a high-quality device like those manufactured by Jensen[20] or Lundahl[21].

In summary, the path to highest analog quality is that which is closest to a straight wire. More is not better; every device removed from the audio path will yield an improvement in clarity, transparency, and fidelity. Use only the minimum number of amplifiers, capacitors, and transformers. For example, never leave a line amplifier or compressor on-line in "test" mode because it seems too much trouble to remove it. Small stations often sound dramatically superior to their "big time" rivals because the small station has a simple audio path, while the big-budget station has put everything but the kitchen sink on-line. The more equipment the station has (or can afford), the more restraint and self-discipline it needs. Keep the audio path simple and clean! Every amplifier, resistor, capacitor, transformer, switch contact, patch-bay contact, etc., is a potential source of audio degradation. Corrosion of patch-bay contacts and switches can be especially troublesome, and the distortion caused by these problems is by no means subtle.

Quality in Digital Chains

In **digital signal processing devices**, the lowest number of **bits per word** necessary to achieve professional quality is 24 bits. This is because there are a number of common

---

[20] Jensen Transformer, Inc., Chatsworth, California, USA (Phone +1 818 374-5857, or Fax +1 818 374-5856)

[21] Lundahl Transformers AB, Tibeliusgatan 7 SE-761 50, Norrtälje SWEDEN (Phone: +46 176 139 30 Fax: +46 176 139 35)

DSP operations (like infinite-impulse-response filtering) that substantially increase the digital noise floor, and 24 bits allows enough headroom to accommodate this without audibly losing quality. (This assumes that the designer is sophisticated enough to use appropriate measures to control noise when particularly difficult filters are used.) If floating-point arithmetic is used, the lowest acceptable word length for professional quality is 32 bits (24-bit mantissa and 8-bit exponent; sometimes called "single-precision").

In **digital distribution systems**, 20-bit words (120dB dynamic range) are usually adequate to represent the signal accurately. 20 bits can retain the full quality of a 16-bit source even after as much as 24dB attenuation by a mixer. There are almost no A/D converters that can achieve more than 20 bits of real accuracy, and many "24-bit" converters have accuracy *considerably* below the 20-bit level. "Marketing bits" in A/D converters are outrageously abused to deceive customers, and, if these A/D converters were consumer products, these bogus claims would be actionable by the Federal Trade Commission.

There is considerable disagreement about the audible benefits (if any) of **raising the sample rate** above 44.1 kHz. An extensive double-blind test[22] using 554 trials showed that inserting a CD-quality A/D/A loop into the output of a high-resolution (SACD) player was undetectable at normal-to-loud listening levels by any of the subjects, on any of four playback systems. The noise of the CD-quality loop was audible only at very elevated levels.

Moreover, there has been at least one rigorous test comparing 48 kHz and 96 kHz sample rates[23]. This test concluded that there is no audible difference between these two sample rates if the 48 kHz rate's anti-aliasing filter is appropriately designed.

However in 2016, a controversial "meta-analysis" of existing tests comparing high-resolution and CD-quality audio was published[24] in the AES Journal. According to the author, "Eighteen published experiments for which sufficient data could be obtained were included, providing a meta-analysis that combined over 400 participants in more than 12,500 trials. Results showed a small but statistically significant ability of test subjects to discriminate high resolution content, and this effect increased dramatically when test subjects received extensive training. This result was verified by a sensitivity analysis exploring different choices for the chosen studies and different analysis approaches. Potential biases in studies, effect of test methodology, experimental design, and choice of stimuli were also investigated. The overall conclu-

---

[22] Meyer, E. Brad; Moran, David R., "Audibility of a CD-Standard A/DA/A Loop Inserted into High-Resolution Audio Playback" JAES Volume 55 Issue 9 pp. 775-779; September 2007

[23] Katz, Bob: *Mastering Audio: the art and the science.* Oxford, Focal Press, 2002, p. 223

[24] Reiss, Joshua D., "A Meta-Analysis of High Resolution Audio Perceptual Evaluation," JAES Volume 64 Issue 6 pp. 364-379; June 2016.

sion is that the perceived fidelity of an audio recording and playback chain can be affected by operating beyond conventional resolution."

Assuming perfect hardware, it can be shown that this debate comes down entirely to the audibility of a given anti-aliasing filter design, as is discussed below. Nevertheless, in a marketing-driven push, the record industry attempted to change the consumer standard from 44.1 kHz to a higher sampling frequency via DVD-A and SACD, neither of which succeeded in the mass marketplace. The industry is trying again with Blu-ray audio and it remains to be seen if they will be more successful than they were with DVD-A or SACD.

Regardless of whether scientifically accurate testing eventually proves that this is audibly beneficial, sampling rates higher than 44.1 kHz have no benefit in FM stereo because the sampling rate of FM stereo is 38 kHz, so the signal must eventually be lowpass-filtered to 17 kHz or less to prevent aliasing. It is beneficial in DAB, which typically has 20 kHz audio bandwidth, but offers no benefit at all in AM, whose bandwidth is no greater than 10 kHz in any country and is often 4.5 kHz.

Some **A/D converters** have built-in soft clippers that start to act when the input signal is 3 – 6 dB below full scale. While these can be useful in mastering work, they have no place in transferring previously mastered recordings (like commercial CD). If the soft clipper in an A/D converter cannot be defeated, that A/D should not be used for transfer work.

Dither is random noise that is added to the signal at approximately the level of the least significant bit. It should be added to the analog signal before the A/D converter, and to any digital signal before its word length is shortened. Its purpose is to linearize the digital system by changing what is, in essence, "crossover distortion" into audibly innocuous random noise. Without dither, any signal falling below the level of the least significant bit will disappear altogether. Dither will randomly move this signal through the threshold of the LSB, rendering it audible (though noisy). Whenever any DSP operation is performed on the signal (particularly decreasing gain), the resulting signal must be re-dithered before the word length is truncated back to the length of the input words. Ordinarily, correct dither is added in the A/D stage of any competent commercial product performing the conversion. However, some products allow the user to turn the dither on or off when truncating the length of a word in the digital domain. If the user chooses to omit adding dither, this should be because the signal in question already contained enough dither noise to make it unnecessary to add more.

Many computer software volume controls do not add dither when they attenuate the signal, thereby introducing low-level truncation distortion. It is wise to bypass computer volume controls wherever possible and if this is not possible, to maintain unity gain through the volume control. Microsoft Windows Media Player and Adobe Flash Players should be operated at 100% 0 dBFS at all times, and level control should be done either at the amplifier volume control or console fader.

In the absence of "**noise shaping**," the spectrum of the usual "triangular-probability-function (TPF)" dither is white (that is, each arithmetic frequency increment contains the same energy). However, noise shaping can change this noise spectrum to concentrate most of the dither energy into the frequency range where the ear is least sensitive. In practice, this means reducing the energy around 4 kHz and raising it above 9 kHz. Doing this can increase the effective resolution of a 16-bit system to almost 19 bits in the crucial midrange area, and is standard in CD mastering. There are many proprietary curves used by various manufacturers for noise shaping, and each has a slightly different sound.

It has been shown that passing noise shaped dither through most classes of signal processing and/or a D/A converter with non-monotonic behavior will destroy the advantages of the noise shaping by "filling in" the frequency areas where the original noise-shaped signal had little energy. The result is usually poorer than if no noise shaping had been used. For this reason, Orban has adopted a conservative approach to noise shaping, recommending so-called "first-order highpass" noise shaping and implementing this in Orban products that allow dither to be added to their digital output streams. First-order highpass noise shaping provides a substantial improvement in resolution over simple white TPF dither, but its total noise power is only 3dB higher than white TPF dither. Therefore, if it is passed through additional signal processing and/or an imperfect D/A converter, there will be little noise penalty by comparison to more aggressive noise shaping schemes.

One of the great benefits of the **digitization of the signal path** in broadcasting is this: Once in digital form, the signal is far less subject to subtle degradation than it would be if it were in analog form, although in fixed point form it is still subject to clipping for reasons discussed earlier in this book. Short of being clipped or becoming entirely un-decodable, the worst that can happen to the signal is deterioration of noise-shaped dither, and/or added jitter.

Jitter is a time-base error. The only jitter than cannot be removed from the signal is jitter that was added in the original analog-to-digital conversion process. *All* subsequent jitter can be completely removed in a sort of "time-base correction" operation, accurately recovering the original signal. The only limitation is the performance of the "time-base correction" circuitry, which requires sophisticated design to reduce added jitter below audibility. This "time-base correction" usually occurs in the digital input receiver, although further stages can be used downstream.

Sample rate converters *can* introduce jitter in the digital domain because they resample the signal, much like A/D converters. Maintaining lowest jitter in a system requires synchronizing all devices in the audio chain to a common wordclock or AES11 signal. This eliminates the need to perform cascaded sample rate conversions on the signals flowing through the facility. Good wordclock generators have very low jitter (also known as "phase noise") and allow the cascaded devices to perform at their best. See *Digital Audio Clock* on page 25.

There are several pervasive myths regarding digital audio:

One myth is that **long reconstruction filters smear the transient response of digital audio**, and that there is thus an advantage to using a reconstruction filter with a short impulse response, even if this means rolling off frequencies above 10 kHz. Several commercial high-end D-to-A converters operate on exactly this mistaken assumption. This is one area of digital audio where intuition is particularly deceptive.

The sole purpose of a reconstruction filter is to fill in the missing pieces between the digital samples. These days, symmetrical finite-impulse-response filters are typically used for this task because they have no phase distortion. The output of such a filter is a weighted sum of the digital samples symmetrically surrounding the point being reconstructed. The more samples that are used, the better and more accurate the result, even if this means that the filter is very long.

It's easiest to justify this assertion in the frequency domain. Provided that the frequencies in the passband and the transition region of the original anti-aliasing filter are entirely within the passband of the reconstruction filter, then the reconstruction filter will act only as a delay line and will pass the audio without distortion. Of course, all practical reconstruction filters have slight frequency response ripples in their passbands, and these can affect the sound by making the amplitude response (but not the phase response) of the "delay line" slightly imperfect. But typically, these ripples are in the order of a few thousandths of a dB in high-quality equipment and are very unlikely to be audible.

The authors have proved this experimentally by simulating such a system and subtracting the output of the reconstruction filter from its input to determine what errors the reconstruction filter introduces. Of course, you have to add a time delay to the input to compensate for the reconstruction filter's delay. The source signal was random noise, applied to a very sharp filter that band-limited the white noise so that its energy was entirely within the passband of the reconstruction filter. We used a very high-quality linear-phase FIR reconstruction filter and ran the simulation in double-precision floating-point arithmetic. The resulting error signal was a minimum of 125 dB below full scale on a sample-by-sample basis, which was comparable to the stopband depth in the experimental reconstruction filter.

We therefore have the paradoxical result that, in a properly designed digital audio system, the *frequency response of the system and its sound is determined by the anti-aliasing filter and not by the reconstruction filter*. Provided that they are realized with high-precision arithmetic, *longer reconstruction filters are always better*.

This means that a rigorous way to test the assumption that high sample rates sound better than low sample rates is to set up a high-sample rate system. Then, without changing any other variable, introduce a filter in the digital domain with the same frequency response as the high-quality anti-aliasing filter that would be required for the lower sample rate. If you cannot detect the presence of this filter in a double-blind test, then you have just proved that the higher sample rate has no intrinsic audible advantage, because you can always make the reconstruction filter audibly transparent.

Another myth is that **digital audio cannot resolve time differences smaller than one sample period and therefore damages the stereo image**.

People who believe this like to imagine an analog step moving in time between two sample points. They argue that there will be no change in the output of the A/D converter until the step crosses one sample point and therefore the time resolution is limited to one sample.

The problem with this argument is that there is no such thing as an infinite-risetime step function in the digital domain. To be properly represented, such a function has to first be applied to an anti-aliasing filter. This filter turns the step into an exponential ramp, which typically has equal pre- and post-ringing. This ramp can be moved far less than one sample period in time and still cause the sample points to change value.

In fact, assuming no jitter and correct dithering, the time resolution of a digital system is the same as an analog system having the same bandwidth and noise floor. Ultimately, the time resolution is determined by the sampling frequency and by the noise floor of the system. As you try to get finer and finer resolution, the measurements will become more and more uncertain due to dither noise. Finally, you will get to the point where noise obscures the signal and your measurement cannot get any finer. However, this point is orders of magnitude smaller in time than one sample period *and is the same as in an analog system.*

A final myth is that **upsampling digital audio to a higher sample frequency will increase audio quality or resolution**. In fact, the original recording at the original sample rate contains all of the information obtainable from that recording. The only thing that raising the sample frequency does is to add ultrasonic images of the original audio around the new sample frequency. In any correctly designed sample rate converter, these are reduced (but never entirely eliminated) by a filter following the upsampler. People who claim to hear differences between "upsampled" audio and the original are either imagining things or hearing coloration caused by the added image frequencies or the frequency response of the upsampler's filter. They are *not* hearing a more accurate reproduction of the original recording.

This also applies to the **sample rate conversion** that often occurs in a digital facility. It is quite possible to create a sample rate converter whose filters are poor enough to make images audible. One should test any sample rate converter, hardware or software, intended for use in professional audio by converting the highest frequency sinewave in the bandpass of the audio being converted, which is typically about 0.45 times the sample frequency. Observe the output of the SRC on a spectrum analyzer or with software containing an FFT analyzer (like Adobe Audition). In a professional-quality SRC, images will be at least 90 dB below the desired signal, and, in SRC's designed to accommodate long word lengths (like 24 bit), images will often be –120 dB or lower, assuming a 24-bit path (which is capable of representing low-level energy down to –144 dBFS). Taking full advantage of high-performance sample rate conversion is another reason to use 24-bit audio for production and to reduce the bit depth (if necessary for applications like burning audio CDs) only as the final step, using appropriate dither.

A good reference on sample rate conversion performance can be found here: http://src.infinitewave.ca/

And finally, some truisms regarding loudness and quality:

Every radio is equipped with a volume control, and every listener knows how to use it. If the listener has access to the volume control, he or she will adjust it to his or her preferred loudness. After said listener does this, the only thing left distinguishing the "sound" of the radio station is its texture, which will be either clean or degraded, depending on the source quality and the audio processing.

Any Program Director who boasts of his station's $20,000 worth of "enhancement" equipment should be first taken to a physician who can clean the wax from his ears, then forced to swear that he is not under the influence of any suspicious substances, and finally placed gently but firmly in front of a high-quality monitor system for a demonstration of the degradation that $20,000 worth of "enhancement" causes! Always remember that less is more.

Content/Ad Insertion/Replacement

Content insertion either replaces audio content in the main program feed and/or supplements it when different program material must be sent to multiple destinations. It may be required because of broadcast rights limitations and/or separate advertising schedules for different destinations. Examples of this include ad insertion for different transmitter feeds and Internet streaming.

Seamless content insertion presents many challenges and operating obstacles. Potential problems include audio miscues, gaps, cut-offs, incorrect rejoin times, and irritatingly inconsistent audio level control, among others. Many of the software and systems that control content insertion have been developed and provided by inexperienced broadcast industry developers. These systems should be chosen carefully so they can provide professional results.

Modern systems use playout system metadata to control content insertion[25]. To ensure correct insertion cue times, the playout system software must deliver on-time, synchronous metadata. Some insertion systems use multiple playout systems and must be controlled from this metadata. If the metadata is used for Internet streaming insertion, it is also important that the streaming protocols and encoders support synchronous metadata. Many of the current streaming protocols currently in use,

---

[25] Some systems, particularly older ones, may use silence sensing or subaudible cue tones in the main program to trigger insertion. Silence sensing introduces gaps of dead air. The filter that removes cue tones can introduce latency that compromises timing, and can increase peak levels, potentially producing clipping distortion on material whose peak level is already close to 0 dBFS. Compared to using metadata, both of these older methods are deprecated.

such as SHOUTcast and Icecast ICY, cannot deliver synchronous metadata, so these systems can never be controlled on-time. Getting insertion timing right is one of many reasons to use StreamS HLSdirect™ Stream Encoders.

There are some insertion systems that require program feeds to be looped through, creating single point of failure: the entire program feed is at the mercy of the insertion system audio capabilities and reliability. If the system fails, the program feed fails.

For Internet streaming, there are systems that address individual players with content targeted specifically to them. Because this is a streaming server-side operation, it can deliver many different simultaneous content insertions to multiple destinations. Although this type of content insertion is valuable because it precisely targets program elements (such as ads) to individual listeners, making it sound professional is very challenging. It cannot perform audio segues and overlays (things that radio does often does to sound polished and professional), so it can compromise production values. Content insertion is done at the network level, not in the main audio program, so the inserted material does not flow through the same audio processing as the main program and can thus sound different from or disconnected from the main program material.

Preventing this requires the content to be precisely loudness-matched and processed to match the main program material. If the audio processing on the main program is changed, then all the insertion content files must be changed to match. Typically, the main program or stream will be processed using Orban Optimod-PC or PCn. The offline file processing functionality of the Orban Optimod-PCn can be used to preprocess the insertion content files.

The Nielsen Audio® PPM (Portable People Meter)

The Nielsen Audio PPM Encoder is an audio watermarking device that adds encoded data about the program audio to the audio itself so that a monitoring device, equipped with a microphone and worn by a listener, can receive the data via acoustic transmission from the radio receiver or computer loudspeaker. The PPM algorithm, which is proprietary to Nielsen, is based on the well-known principle of psychoacoustic masking. For most listeners, the program material masks or "drowns out" the added data in to render it inaudible.

To maximize the data throughput, the average level of the program audio should be consistently high. This maximizes the ability of the PPM encoder to inject its data while ensuring that the program audio masks the data. While a simple AGC will help compared to no audio processing at all, a full audio processing chain including an AGC, multiband compressor, and peak limiting will work significantly better than an AGC alone.

Because the PPM signal amplitude is very low with respect to the program audio, the PPM signal can, in principle, be added to the final peak limited audio signal without significantly disturbing peak modulation and without compromising loudness. However, there are potential pitfalls. For digital broadcasting or netcasting, adding the PPM signal is easily done by inserting the Nielsen PPM encoder after the peak limited signal just before the broadcast/netcast digital audio encoder input. This is best done in the digital domain by using an Nielsen PPM digital encoder, which does not compromise the program audio waveform fidelity and which has sufficient headroom. If a Nielsen PPM analog encoder is used, one must to pay careful attention to headroom to prevent the encoder from clipping the audio. Furthermore, it is necessary to determine if the analog encoder signal path introduces overshoot and/or tilt into the processed audio. If it does so, it can cause peak clipping to occur in the transmission chain after the analog encoder. Correcting this requires the gain after the encoder to be lowered, which reduces the loudness of the transmission, and partially defeats the purpose of the audio processing system.

The PPM encoder uses an elegant psychoacoustic audio masking model to combine digital data, called a "watermark," with the actual audio signal. The same principle is used in perceptual audio codecs. By capturing, measuring, processing, and analyzing signals from a Digital Nielsen PPM encoded broadcast program line (without attempting to reverse-engineer the bitstream data format), we were able to determine the following details:

- There are 10 frequency bins, which are located between 1 and 3kHz.
- PPM decoders are very pitch-sensitive. This is very important to consider when using PPM encoded signals for netcasting using Adobe Flash Players. Numerical inaccuracies in the sample rate converter of the Adobe Flash encoder and player render the PPM signal useless at sample rates other than 44.1 kHz. This is unfortunate because a 32 kHz sample rate optimizes the performance of HE-AACv2 low-bitrate netcasting.
- The encoded output signal is the ratio of the input audio signal and the PPM signal.
- The Digital PPM encoder has a five-sample delay at 44.1 kHz: 113.38us.
- The Digital PPM Low Level Audio Alarm threshold is –18dBFS using normalized pink noise or –20dBFS using a 1 kHz sinewave.
- In installations where the program line has 18-20dB of headroom (typical of good engineering practice), the Low Level Audio Alarm may often be in the Alarm state.
- The Digital PPM Low Level Audio Alarm has somewhat inconsistent behavior: It takes as long as 30 to 60 seconds to reset from a Low Signal Level condition, from no audio to normal level audio, and takes as long as 10 minutes to enter Low Signal Level condition after a transition occurs between normal audio level and no audio.

- The Digital PPM encoder continues to encode under Low Signal Level alarm conditions.
- The alarm is independent of PPM encoder/injection functions.

When audio dynamic range compression (not to be confused with bit reduction compression) is used after the PPM encoder, the low level audio alarm appears to be unimportant because the compression will maintain both modulation levels and PPM signal levels.

The encoder PPM signal injection is linearly proportional to the audio level at the encoder's input. Therefore, the location of the PPM encoder (before or after the dynamic range compressor) does not materially affect data throughput. If the compressor is placed after the PPM encoder (as it is in most facilities), the audio signal processing systems will amplify the PPM signal and program audio equally, so the location will not affect the PPM encoded signal except for a slight, inaudible signal-to-noise penalty. However, placing the compressor before the PPM encoder will prevent the low level audio alarm from tripping because the compression keeps the level constant at the input of the PPM encoder.

While the PPM system attempts to increase ratings accuracy by replacing the old paper diary reporting system with an automated one, there are a number of technical reasons why this is not completely foolproof. High average audio levels are required to maintain PPM signal integrity. Program material with many quiet periods such as talk formats offers less opportunity to produce an encoded signal. This is a limitation of this kind of technology.

It is clear that using considerable amounts of audio dynamic range reduction can help maximize PPM data throughput and give stations higher ratings. Orban Optimods can provide high average audio levels for PPM encoders without introducing side effects that drive audiences away due to listening fatigue. A PPM explanation and performance plots are available here:
http://www.indexcom.com/ppm

As of this writing, Nielsen is supplying clients with second-generation encoders. These were introduced around 2016 in response to the introduction of a third-party device that amplifies the Nielsen watermark to make it easier to decode at the expense of potential watermark audibility. The second-generation Nielsen encoder creates a watermark at the highest possible level without its being audible to Nielsen's panel of "golden-eared" listeners. We believe it is unwise to use a third-party device to further increase the level of the watermark, as it is very likely to be audible and annoying to listeners.

Part 3: Configuring and Using the Production Studio

The role of the production studio varies widely from station to station. If used only for creation of spots, promos, IDs, etc., production studio quality is considerably less critical than it is where programming is "sweetened" before being transferred to a playout system. Our discussion focuses on the latter case, but discusses both.

Monitor Loudspeakers

Choosing Monitor Loudspeakers

The loudspeakers are the single most important influence on studio quality. The production studio monitor system is the quality reference for all production work, and thus for the final sound to be broadcast/netcast. Achieving accurate monitor sound requires considerable care in the choice of equipment and in its adjustment.

Loudspeakers should be chosen to complement room acoustics. The space limitations in production studios usually dictate the use of bookshelf-sized speakers. You should assess the effect of equalization or other sweetening on small speakers to make sure that excessive bass or high-frequency boost has not been introduced. While such equalization errors can sound spectacular on big, wide-range speakers, it can make small speakers with limited frequency response and power-handling capacity sound terrible. The Auratone Model 5C Super Sound Cube has frequently been used as a small speaker reference. Although these speakers are no longer manufactured, they are often available on the used market. We recommend that every production studio be equipped with a pair of these speakers or something similar and that they be regularly used to assure the production operator that his or her work will sound good on small table and car radios.

The primary monitor loudspeakers should be chosen for:

- high power-handling capacity
- low distortion
- high reliability and long-term stability
- controlled dispersion (omnidirectional speakers are *not* recommended)
- good tone burst response at all frequencies
- lack of cabinet diffraction
- relatively flat axial and omnidirectional frequency response from 40-15,000Hz
- physical alignment of drivers (when all drivers are excited simultaneously, the resulting waveforms should arrive at the listener's ears simultaneously, sometimes called "time alignment").

There are a number of powered midfield monitors available from a large assortment of pro-audio companies, like JBL, KRK, Focal, Mackie, Genelec, Tannoy, and Alesis, among others. These speakers are very convenient to use because they have built-in power amplifiers and equalizers. Because they have been designed as a system, they are more likely to be accurate than random combinations of power amplifiers, equalizers, and passive loudspeakers. They are also less likely to be connected out of phase, which will cause dramatic loss of low frequencies, vague stereo imaging, and an overall hollow sound. The principal influence on the accuracy of these powered speakers (particularly at low frequencies) is room acoustics and where the speakers are placed in the room. Some of these speakers allow the user to set the bass equalization to match the speaker's location. We believe that such speakers are a logical choice for main monitors in a broadcast production studio.

Loudspeaker Location and Room Acoustics

The bass response of the speakers is strongly affected by their location in the room. Bass is weakest when the speaker is mounted in free air, away from any walls; bass is most pronounced when the speaker is mounted in a corner. Corner mounting should be avoided because it tends to excite standing waves. The best location is probably against a wall at least 18 inches (45 cm) from any junction of walls. If the bass response is weak at this location because the speaker was designed for wall-junction mounting, it can be corrected by equalization (discussed below). It is important that the loudspeakers be located to avoid acoustic feedback into the turntable, because this can produce a severe loss of definition (a muddy sound).

Many successful monitoring environments have been designed according to the "LiveEnd/Dead-End" (LEDE™) concept invented by Don Davis of Synergistic Audio Concepts. Very briefly, LEDE-type environments control the time delay between the arrival of the direct sound at the listener's ear and the arrival of the first reflections from the room or its furnishings. The delay is engineered to be about 20 milliseconds. This usually requires that the end of the room at which the speakers are mounted be treated with a sound-absorbing material like Sonex® so that essentially no reflections can occur between the speakers' output and the walls they are mounted on or near. Listeners must sit far enough from any reflective surface to ensure that the difference between the distance from the speaker to the listener and the distance from the speaker to the reflective surface and back to the listener is at least 20 feet (6 meters). It is also desirable that the reflections delayed more than 20 milliseconds be well-diffused (that is, with no flutter echoes). Flutter echoes are usually caused by back-and-forth reflections between two parallel walls, and can often by treated by applying Sonex or other absorbing material to one wall. In addition, "quadratic residue diffusors" (manufactured by RPG Diffusor Systems, Inc.) can be added to the room to improve diffusion and to break up flutter echoes.

An excellent short introduction to the theory and practice of LEDE design is Don Davis's article, "The LEDE Concept" in Audio Vol.71 (Aug. 1987): p.48-58. (For a more definitive discussion, see Don and Carolyn Davis, "The LEDE Concept for the Control of Acoustic and Psychoacoustic Parameters in Recording Control Rooms." *J. Audio Eng. Soc.* Vol.28 (Sept. 1980): p.585-95.)

It should be noted that the LEDE technique is by no means the only way to create a good-sounding listening environment (although it is perhaps the best-documented, and has certainly achieved what must be described as a quasi-theological mystique amongst some of its proponents). Examples of other approaches are found in the August 1987 (vol. 29, no. 8), issue of Studio Sound, which focused on studio design.

Loudspeaker Equalization

The performance of any loudspeaker is *strongly* influenced by its mounting location and room acoustics. If room *acoustics are good,* the third-octave real-time analyzer provides an extremely useful means of measuring any frequency response problems intrinsic to the loudspeaker, and of partially indicating problems due to loudspeaker placement and room acoustics.

By their nature, the third-octave measurements combine the effects of direct and reflected sound. This may be misleading if room acoustics are unfavorable. Problems can include severe standing waves, a reverberation time which is not well-behaved as a function of frequency, an insufficient number of "normal modes" (Eigenmodes), lack of physical symmetry, and numerous problems which are discussed in more detail in books devoted to loudspeakers and loudspeaker equalization.

Time-Delay Spectrometry" (TDS) is a technique of measuring the loudspeaker/room interface that provides much more information about acoustic problems than does the third-octave real-time analyzer. TDS (which some sound contractors are licensed to practice) is primarily used for tuning recording studio control rooms, and for adjusting large sound reinforcement systems. The cost may be prohibitive for a small or medium-sized station, particularly if measurements reveal that acoustics can only be improved by major modifications to the room. However, TDS measurements are highly useful in determining if LEDE criteria are met, and will usually suggest ways by which relatively inexpensive acoustic treatment (absorption and diffusion) can improve room acoustics.

With the advent of low-cost personal computers and sound cards, it is possible to buy economical software to do room analysis and tuning. Since the invention of TDS, a number of other techniques like MLSSA (Maximum-Length Sequence System Analyzer; http://mlssa.com) have been developed for measuring and tuning rooms with accuracy greater than that provided by traditional third-octave analyzers.

It is certainly true that room acoustics must be optimized as far as economically and physically possible *before* electronic equalization is applied to the monitor system. (If room acoustics and the monitor are good, equalization may not be necessary.)

Once room acoustic problems have been solved to whatever extent practical, make frequency response measurements to determine what equalization is required. A MLSSA analyzer, a TDS analyzer, dual-channel FFT analyzer, or a third-octave analyzer can be used for the measurements. To obtain meaningful results from the analyzer, the calibrated microphone that comes with the analyzer should be placed where the production engineer's ears would ordinarily be located. If a third-octave analyzer is used, excite each loudspeaker in turn with pink noise while observing the acoustic response on the analyzer. If a MLSSA or TDS analyzer is used, follow the manufacturer's instructions.

Place the analyzer test microphone about 1m from the monitor speaker. Adjust the equalizer (see its operating manual for instructions) to obtain a real-time analyzer read-out that is flat to 5 kHz, and that rolls off at 3dB/octave thereafter. (A truly flat response is not employed in typical loudspeakers, and will make most recordings sound unnaturally bright and noisy.)

Electronic equalization cannot fix acoustic nulls in the room caused by standing waves. Nulls should be corrected by acoustic treatment of the room and by careful placement of the loudspeakers. A good rule of thumb is never to set an equalizer to create a large, narrowband boost because this added energy will probably sound unnatural elsewhere in the room where the null does not exist.

If the two channels of the equalizer must be adjusted differently to obtain the desired response from the left and right channels, suspect room acoustic problems or poorly matched loudspeakers. The match is easy to check: just physically substitute one loudspeaker for the other, and see if the analyzer reads the same. Move the microphone over a space of two feet or so while watching the analyzer to see how much the response changes. If the change is significant, then room acoustic problems or very poorly controlled loudspeaker dispersion is likely. If it is not possible to correct the acoustic problem or loudspeaker mismatch directly, you should at least measure the response at several positions and average the results. (Microphone multiplexers can automatically average the outputs of several microphones in a phase-insensitive way—they will help you equalize loudspeaker response properly.)

Although left and right equalizers can be adjusted differently below 200Hz, they should be set close to identically above 200Hz to preserve stereo imaging, even if this results in less than ideal curves as indicated by the third-octave analyzer. (This is a limitation of the third-octave analyzer, which cannot distinguish between direct sound, early reflections, and the reverberant field; stereo imaging is primarily determined by the direct sound.)

A few companies are now making DSP-based room equalizers that attempt to correct both the magnitude and phase of the overall frequency response in the room. These can produce excellent results if the room is otherwise acoustically well behaved.

Recently, several companies[26] have developed room correction equalizers that rely on several measurements at different locations in the room. They claim that their software can process the results of the multiple measurements to avoid equalizing localized acoustic anomalies.

[26] For example, http://www.audyssey.com/ and https://www.trinnov.com/

Finally, we note once again that the manufacturers of powered nearfield monitors have done much of the work for you. These monitors have built-in equalization, which will often be quite adequate even at low frequencies if the monitor's equalizer can be set to complement the monitor's location in the room.

Other Production Equipment

The discussions in this document of disk reproduction, tape, digital source, and electronic quality also apply to the production studio. Uncompressed sources, including CD, DVD-A, SACD, and losslessly compressed files usually provide the highest quality. For cuts that must be taken from vinyl disk, it is preferable to use "high-end" consumer phono cartridges, arms, and turntables in production, and be sure that the turntable's speed is exactly correct. (This is customarily done with a test record having a stobe pattern on the label.) Make sure that *one* person has responsibility for production quality and for preventing abuse of the record playing equipment. Having a single production director will also help achieve a consistent air sound—an important contribution to the "big-time" sound many stations want.

There are many low-cost all-digital mixers available. Made by companies like Soundcraft, Yamaha, Mackie, and Roland, these provide the ability to automate mixes and to keep the signal in the digital domain throughout the production process.

Although some people still swear by certain "classic" vacuum-tube power amplifiers (notably those manufactured by Marantz and McIntosh), the best choice for a monitor amplifier is probably a medium-power (100 watts or so per channel) solid-state amplifier with a good record of reliability in professional applications.

Production Practices

The following represents some of our opinions on production practices. We are aware that some production facilities operate under substantially different philosophies. But we feel that the recommendations below are rational and offer a good guide to achieving consistently high quality.

1. **Do not apply general audio processing to dubs and syndicated programs from commercial recordings in the production studio.**

   Optimod provides all the processing necessary, and does so in real time with a remarkable lack of audible side effects. Further compression is not only undesirable but is likely to be very audible. If the production compressor has a slow attack time (and therefore produces overshoots that can activate gain reduction in Optimod), it will probably "fight" with a downstream Optimod, ultimately yielding a substantially worse air sound than one might expect given the individual sounds of the two units. Because real-time Optimod processing takes into account transitions between program elements, it can handle these more smoothly than file-based processing that processes each file in the playout without regard to context.

If it proves impossible to train production personnel to record with the correct levels, we recommend using the Orban Optimod-PCn to protect the production recorder from overload. When used for leveling only, Optimod-PCn does not affect short-term peak-to-average ratio of the audio, and so will not introduce unnatural artifacts into Optimod processing. Optimod-PCn is a pure software audio processor and can be used in Windows or higher computers with Intel i-series CPUs, such as the one that may already be present in the production studio.

2. **Do not modify what you cannot hear.**

   Noise-induced and/or age-related hearing loss is common for those working in the audio industry and it is often emotionally difficult for practitioners to accept that they no longer hear as well as they once did. It is wise for those applying processing in the production studio to take a hearing test to determine if there are parts of the audio spectrum that are no longer audible to them. If so, they should avoid applying equalization or other treatment in spectral regions that cannot hear. Moreover, significant high frequency loss makes it impossible to fully assess the effects of single-ended noise reduction.

3. **Avoid excessive bass and treble boost.**

   Substandard recordings can be sweetened with equalization to achieve a tonal balance typical of the best currently produced recordings. However, avoid excessive treble boost because it will stress AM and FM audio processors. We recommend using a modern CD typical of your program material as a reference for spectral balance although not for dynamics processing because of the excessive limiting and clipping applied to all too many of today's CDs. Very experienced engineers master major-label CDs using the best available processing and monitoring equipment, typically costing over $100,000 per room in a well-equipped mastering studio. The sound of major-label CDs represents an artful compromise between the demands of different types of playback systems and is designed to sound good on all of them, although this goal is often compromised by today's CD loudness wars. Mastering engineers do not make these compromises lightly. We believe it is very unwise for a radio station to significantly depart from the spectral balance typical of major-label CDs because this almost certainly guarantees that there will be a class of receivers or players on which the audio sounds terrible.

4. **Pay particular attention to the maintenance of production studio equipment.**

   Even greater care than that employed in maintaining broadcast equipment is necessary in the production studio because quality loss here will repeatedly appear on the air. The production director should be acutely sensitized to audible quality degradation and should immediately inform the engineering staff of any problems detected by ear.

5. **Minimize motor noise.**

   To prevent motor noise from leaking into the production microphone, equipment with noisy fans and hard drives should be installed outside the studio if possible. Otherwise, they should reside in alcoves under soffits, surrounded by acoustic treatment. In the real world of budget limitations this is sometimes not possible, although sound-deadening treatment of small spaces is so inexpensive that there is little excuse for not doing it. But even in an untreated room, it is possible to use a directional microphone (with figure-eight configuration, for example) with the noisy machine placed on the microphone's "dead" axis. Choosing the frequency response of the microphone to avoid exaggerating low frequencies will help. In particularly difficult cases, a noise gate or expander can be used after the microphone preamp to shut off the microphone except during actual speech.

6. **Consider processing the microphone signal.**

   Audio processing can be applied to the microphone channel to give the sound more punch. Suitable equalization may include gentle low- and high-frequency boosts to crispen the sound, aid intelligibility, and add a "big-time" quality to the announcer. But be careful not to use too much bass boost, because it can *degrade* intelligibility. Effects like "telephone" and "small transistor radio" can be achieved with equalization, too.

   The punch of production material can often be enhanced by tasteful application of compression to the microphone chain. However, avoid using an excessive amount of gain reduction and excessively fast release time. These cause room noise and announcer breath sounds to be exaggerated to grotesque levels (although this problem can be minimized if the compressor has a built-in expander or noise gate function).

   When adjusting the microphone processor, adjust the main audio processor for your desired sound on music first and then adjust the microphone processor to complement the main processing you have selected.

   Close micing, which is customary in the production studio, can exaggerate voice sibilance. In addition, many women's voices are sibilant enough to cause unpleasant effects. High-frequency equalization and/or compression will further exaggerate sibilance. If you prefer an uncompressed sound for production work but still have a sibilance problem, then consider locating a dedicated de-esser *after* all other processing in the microphone chain. It might be necessary to use personalized microphone processing settings for different announcers, particularly if some are male and some are female. A secret weapon for a very distinctive sound can be a multiband microphone processor.

7. **Use single-ended noise reduction and de-clipping software with care.**

   Single-ended noise reduction can cause objectionable artifacts, and de-clipping software can sometimes made distortion worse, not better. See *Restoration Software* starting on page 71. In general, you should always archive a copy of the raw audio data prior to restoration because restoration software is likely to improve in the future.

8. **Beware of listening fatigue.**

   Listening fatigue sets in after about 30 minutes[27] of concentrated listening. Fatigue causes people to become less sensitive to the effect of changes and makes it more and more difficult to judge whether a change is beneficial, so limit the duration of adjustment sessions.

9. **Do not project your preferences onto an audience without research.**

   One of the hardest lessons for audio professionals to learn is that different people have difference tastes and preferences: one person's "bad" might be another person's "good," and that's OK. It's just how humans are, and psychologically projecting one's personal preferences onto an audience can lead to surprises. The best processing is processing that maximizes your overall audience share, and this requires research and discipline.

Quality Control in Transfers to Playout Systems

Quality control in transfers from sources to playout systems requires attention to detail. In addition to technical problems with the transfer itself, there can also be various problems with sources. This section discusses both.

Program Material Quality and Authenticity

Many original record labels are defunct and have transferred licenses to other labels. It was formerly safe to assume that the audio from the original record/CD label or authorized licensee is as good as it gets, but tasteless remastering has ruined many recent major label re-releases, even within the same labels. Moreover, at this writing at least one major music conglomerate that is universally familiar is known to apply watermarking codes that are sometimes audible, particularly with music having wide dynamic range, and to do so even on files marketed as "high resolution." In this case, it is worthwhile to seek out older sources of the same material without watermarking.

Many major labels produce collections for other well-known marketing groups. Many of these sources are acceptable, although they require careful auditioning and quality validation. Some smaller and obscure labels have acquired licenses from the original labels. While some of this work has proven to be excellent, some of these reissues should be avoided.

---

[27] Recommendation ITU-R BS.1116-1: "Methods for the Subjective Assessment of Small Impairments in Audio Systems including Multichannel Sound Systems," section 4.2. While specifically applicable to extremely critical listening tests, this document contains much generally useful information about setting up reliable listening tests.

Syndicated programming can be another source of audio quality problems, especially if the audio source material has been preprocessed in any way. Such programming should be very carefully validated.

Some music was poorly recorded and mastered. There are many reasons for this, including poor equipment or monitoring environments. Some cases will require careful re-equalization to achieve consistent results.

Many tracks, even from "desirable" labels, have been recently re-mastered and may sound quite different from the original transfer to CD. Regardless of source, it is wise to use the original performance even if its audio quality is worse than alternative versions. Sometimes the original performance has been remixed for digital media release, which often improves the quality. However, beware of remixes so radical that they no longer sound like the hit version as remembered by audiences. Furthermore, many of these original performances may be available on many different albums, possibly from the same label but from different sources with varying quality. A valuable on-line resource for this information is *Top 40 Music on Compact Disc*: http://www.top40musiconcd.com/

Mono mixes often sound punchier than stereo mixes, and oldies formats specializing in music earlier than 1970 play a plethora of material that audiences may remember from the original mono 45 rpm vinyl. This authenticity is an argument for playing the original mono mixes in such formats.

However, recent advances in digital signal processing algorithms and software have made *Digitally-Extracted Stereo* (DES) possible. Audio spectral editing, sound source separation, and other related tools and processes are now being used specifically for the purpose of upmixing mono source material to stereo, with the goal of creating stereo mixes that are virtually indistinguishable from stereo mixes created using multitrack session tapes, had they existed[28]. At this writing, this still requires substantial manual effort and its success or failure depends on the taste and "ears" of the person doing the extraction, as well as the suitability of the material. The procedure starts with source separation to extract various elements, like vocals, percussion, guitars etc. on independent, synchronized tracks. These can then be mixed like other multitrack sources, paying close attention to preserving mono compatibility and to balancing the mix so it is faithful to the original performance.

Some of the more recent remasterings may contain additional signal processing beyond simple click and pop elimination. Newly remastered tracks should be validated very carefully, as the newer tracks may suffer from excessive digital limiting that reduces transient impact and punch. Therefore, the older, less-processed sources may stand up better to Optimod transmission processing.

[28] https://www.monotostereo.info/

Unfortunately, there is no reliable formula for choosing old or new CDs. For example, some original CD releases were simply transfers of a vinyl pressing. The limitations of vinyl are usually audible and subsequent remasters were a dramatic improvement, either because the original master tapes were discovered and used or because improved vinyl restoration software and techniques were employed. On the other hand, many remasters were subject to additional dynamic range compression and peak limiting and do not sound as good as their original releases even though the newer remasters may claim higher resolution.

We believe that the best-sounding CDs are probably those mastered from about 1990 to 1995. Before 1990, many mastering engineers used the Sony PCM1610's converters because the standard medium for transmitting a mastered CD to a replication house was then a 3/4" "U-Matic" video recorder and the Sony formatted the audio so that it could be recorded as video on such a machine. However, the PCM1610's converters were widely criticized for their sound. There were too many low-slewrate opamps and electrolytic capacitors in the signal path, not to mention high-order analog anti-aliasing filters.

Things changed around 1990 with the introduction of analog to digital converters based on the then-new Analog Devices AD1879 delta/sigma A/D converter chip. This chip uses oversampling to eliminate the need for high-order analog anti-aliasing filters and has very good low-level linearity. The resulting signal path has very little group delay distortion. If properly designed, converters based on the AD1879 sound very good. Meanwhile, the loudness wars were still at least five years in the future.

Starting around 1995, average levels on CDs started to increase. The availability of early digital look-ahead limiters like the WAVES L1 enabled CD mastering engineers to limit peaks without obvious side effects. However, like most anything else, the availability of the tools led to their abuse. Digital limiting, which was a bit like crack to some mastering engineers, started to suck the life and punch out of material for the sake of loudness. Certain engineers developed a reputation as "go-to guys" if you wanted a "loud" CD. Look-ahead limiting by itself was no longer enough — some engineers started to run material through analog clippers prior to A/D conversion. This allowed them to chop off peaks caused by snare drums and the like without the pumping that digital limiters could add to such material.

Even that wasn't enough. Because the life was sucked out of the material by too much clipping and limiting, some engineers started tarting up the corpse with yet more signal processing before the peak limiters. For examples, vacuum tube equalizers and compressors were used to add some sparkle by driving the tubes into soft distortion. While this could actually help some material that was too sterile and that "didn't sound like a record," once again it could be and was abused.

All too often, today's digital tracks are squashed into fatiguing mush by overprocessing. Fingernails-on-a-chalkboard brightness combined with dynamic flatness to create sound that many music consumers find disturbing without really knowing why.

Although the radio broadcast community has used processing before transmission for most of the history of the medium, this has been thoroughly researched over the years to discover how it affects audiences. In general, the type of processing used on typical digital sources these days has been shown to increase “cume” (the number of distinct persons listening to a given station or stream in a one-week period) while driving average time spent listening down. Moreover, the kind of brightness present on many of the today’s digital sources has been shown by broadcasters to repel women listeners. It is probably no coincidence that most mastering engineers are male! The theory “If a little sounds good, a lot must be better,” usually *does not* apply to audio signal processing.

Meanwhile, as CDs became more and more overprocessed, their sales declined precipitously. We suspect that streaming was not the only cause. We have to wonder why the executives who run the labels refuse to make the connection between the sort of brutal overprocessing on many of today’s CDs and the increasing lack of satisfaction with the product. It’s not as if they haven’t been told!

Here are some interesting references on the CD Loudness Wars:
http://www.npr.org/templates/story/story.php?storyId=122114058
http://flowingdata.com/2010/01/05/a-visual-history-of-loudness-in-popular-music/

It is simplistic to complain about loud CDs based solely on their BS.1770 integrated loudness without listening and taking into account how the loudness was created. An unfortunate aspect of the CD loudness wars is that for a given loudness, most CDs do not sound as good as they could if optimum processing was used in production and mastering. It is possible to create big, contemporary-sounding masters without the fatiguing distortion, overt clipping, grating high frequencies, and dynamic deadness of so much of today’s ”hypercompressed” commercially released music. Orban Optimod-PCn 1600 software for Windows uses refined technology originally developed for broadcast signal processing to allow very loud audio files to be created without objectionable artifacts.

Vinyl

While we had expected the black vinyl disk to be obsolete by this revision, it is still used in specialized applications like live “club-style” D.J. mixing and re-mastering when the original recordings are no longer available or are in poor condition.

Moreover, vinyl is enjoying a modest revival as a consumer music source. Part of its appeal is that it is typically mastered more conservatively than “hypercompressed,” “loudness-at-all-costs” digital media, and vinyl can serve broadcasters/netcasters as a better-sounding alternative source when the only available digital sources have been abused in mastering. However, beware of current vinyl releases that have been cut from “hypercompressed” CD masters; in this case, the damage has already been done.

If the vinyl is in less than pristine condition, it can often be improved by restoration software. Transfers from vinyl are discussed in detail in *Vinyl Disk* starting on page 81.

Mechanics of Transfer

If a linear PCM or losslessly-compressed digital audio file is unavailable from the label or syndicator, by far the best way to ingest audio into playout systems is to extract the digital audio from the CD using a computer and a program to "rip" the audio tracks to a digital file. When done correctly, this has the absolute lowest risk of reducing source audio quality. Recording audio from an external playback device into a computer risks reducing audio quality, as many things can go wrong, including gain changes and 0 dBFS+ overloads.

To achieve the best accuracy when playing out from an external CD player, use one with a digital output (SPDIF, standardized as IEC 60958 type II) between the CD player and the digital playout system. If a CD cannot be ripped accurately in a computer, you will sometimes achieve usable results by playing out the CD in a stand-alone player and recording from the SPDIF output, as stand-alone players provide often-undetectable interpolation to conceal errors that cannot be fully corrected. In its reviews of disk players, Stereophile Magazine[29] reports error correction and interpolation abilities. This is typically done by using a Pierre Verany test CD titled *Compact Test Demonstrations*, which has precisely calibrated gaps of various sizes in the data.

The primary advantage of computer ripping is speed. However, it is crucial to use the right hardware and software to achieve error correction equivalent to that routinely found in a stand-alone CD player. A combination of an accurate extraction program (such as Exact Audio Copy or EAC, which has best metadata handling; http://www.exactaudiocopy.de/, or CUERipper, an open-source alternative to EAC) and a Plextor® duplicator drive[30] (which implements hardware error correction) will yield exceptional results and will automatically log and detect uncorrectable errors. Not all drives are capable of digital audio extraction and not all drives offer hardware error correction. Make sure you drive supports "Accurate Stream" and C2 error reporting; this will guide the ripping software into making multiple passes to try to overcome read errors, and will allow the software to report its ultimate success or failure in making an error-free rip. Moreover, bear in mind that once the ripped audio file is passed through other digital processing, it is subject to the 0 dBFS+ issue unless said processing has been designed to allow sufficient headroom.

The CD CHECK Test Disc is useful to validate a CD player or CD drive error correction and is available from Digital Recording at: :
http://www.digital-recordings.com/cdcheck/cdcheck.html

---

[29] https://www.stereophile.com/

[30] As of this writing, the recommended Plextor model is PX-891SAF-Plus.

It is also possible to extract or rip DVD-A audio if multichannel audio is required. A very capable application is DVD Audio Extractor and available here: http://www.dvdae.com/

As of this writing, SACD DSD cannot legally be digitally extracted or ripped. SACD Hybrid discs with a CD layer can be ripped using the standard CD digital audio extraction method described earlier. If the SACD layer has been mastered without the quality-degrading peak limiting and clipping that have become ubiquitous on contemporary CD releases, an analog transfer of the SACD layer will provide the highest audio quality available.

If you cannot successfully rip a given track without errors by the methods described above, try using restoration software to conceal clicks or discontinuities. See *Restoration Software* on page 71.

Analog Connections

A stand-alone CD player can be used to transfer audio for a digital playout system. The player's analog outputs are subject to 0 dBFS+ overloads, as explained below. The primary advantages to real-time analog transfers are that the transfer engineer can detect any audio glitches caused by damaged or defective CDs and that is that it is easy to control levels during the transfer. This includes increasing the level of song intros, which is a common practice. If you do not need to adjust levels, it is best to use the player's digital output (if available) to make a pure digital connection to the record device. Even if you do have to adjust levels, it is best to do a digital transfer and then adjust the levels in editing software after the transfer has occurred.

An analog connection from an HDCD-enabled CD player is the only way to transfer HDCD-encoded material in a way that preserves all of the HDCD processing, including the program-adaptive filtering. (See *HDCD* on page 8.)

Analog Operating Levels

There are two common operating levels found in analog equipment; +4 dBu for professional equipment and –10 dBV for consumer and "semi-pro" equipment. Mixing the two in a given system can cause clipping and/or loss of signal-to-noise ratio. See *Headroom* on page 31.

0 dBFS+ (True-Peak Overloads)

A peak overload issue commonly called "0 dBFS+" or "true-peak clipping" can be a problem with the analog outputs of source players and computer sound cards, or if sample rate conversion is applied to a bit-accurate digital rip. This will not occur with a bit-accurate digital rip, and is another reason to prefer digital rips or lossless file downloads to analog connections. If the DAC/reconstruction filter does not have 3dB of headroom above 0 dBFS, the DAC is likely to clip the audio coming from today's aggressively hyper-processed digital sources. This will add even more distortion to the regrettable amounts of clipping distortion that are already introduced in the mastering process unless oversampled limiting (which anticipates and compensates for 0 dBFS+ overshoots) was used when the source was mastered. However, many of today's sources are hard-clipped in the digital domain and are therefore likely to excite the 0 dBFS+ phenomenon.[31]

Figure 2 on page 66 shows an example of a sampled waveform whose true peak level is 3 dB higher than its highest digital sample value. The waveform is a sinewave at exactly one-quarter the sampling frequency, and it is sampled at the sinewave's 45° points, which are shown as orange dots at 0 dBFS (which equals ±1 on the graph's linear Y-axis). The curve passing through the orange dots shows the sinewave after D/A conversion, assuming that the D/A converter has headroom above 0 dBFS. The red dots show the peak value's being ±1.41 (+3 dBFSTP, where "TP" denotes "true-peak"). To prevent this waveform from exceeding 0 dBFSTP, its amplitude has to be reduced by 3 dB, such that the samples (blue dots) are now at ±0.707. Then the waveform drawn through the blue samples hits exactly ±1 (0 dBFSTP).

This is a serious issue in the design of D/A converters. A few converters allow headroom above 0 dBFSTP in their analog signal paths, but the vast majority will simply clip off any material above 0 dBFS. This is the most likely explanation for audible differences between different D/A converter designs. We know of at least one manufacturer who takes this into account in the design of its D/A converters.[32]

---

[31] See https://www.indexcom.com/whitepaper/zerodbfsplus/

[32] https://benchmarkmedia.com/blogs/application_notes/intersample-overs-in-cd-recordings. "Every D/A chip and SRC chip that we have tested here at Benchmark has an intersample clipping problem! To the best of our knowledge, no chip manufacturer has adequately addressed this problem. For this reason, virtually every audio device on the market has an intersample overload problem. This problem is most noticeable when playing 44.1 kHz sample rates." John Siau, February 10, 2017.

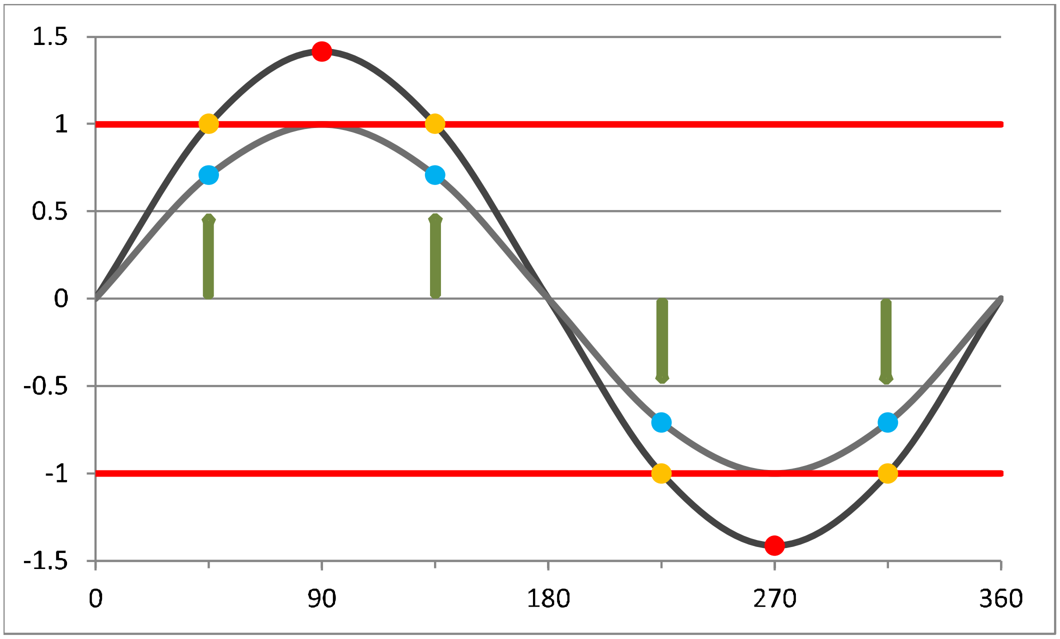

Figure 2: Sinewave at 0.25 fs sampled at 45 degree points, showing +3 dBFSTP true peak level

Conventional wisdom holds that pure digital connections from a CD player or a digital-domain rip cannot cause headroom problems. However, 0 dBFS+ can also be a problem in the digital domain. Passing a digitally-clipped signal through a sample rate converter, even one whose output sample rate is the nominally the same as the input sample rate, can cause overshoots because the SRC interpolates samples between those existing at the input and the interpolated samples can have a higher level than the input samples. Therefore, even digital signal paths can cause the 0 dBFS+ problem, so competently designed digital systems must have enough headroom to prevent clipping in the digital domain caused by 0 dBFS+.

When using computer sound card analog outputs, it is a good idea to make sure the audio levels are no higher than –3dBFS. This means that when you rip CDs into a playout system that uses the analog outputs of a sound card, you should reduce the level of the audio by 3 dB. Before you do this, it is important to verify that the DSP implementing the level change adds appropriate dither (see page 44). If the bit depth is held constant, failure to add dither before a gain reduction in the digital domain will introduce distortion. If the gain reduction occurs in the analog domain, this should not be a problem because any properly designed analog-to-digital converter following the analog gain adjustment will add dither as necessary.

Screening Sources for Technical Problems

When one builds a music library on a digital delivery system, it is important to screen and validate all audio sources. A track's being available on CD does not guarantee good audio quality; some library providers supply .wav files that have been converted from MP3 sources. There are several applications available that attempt to identify .wav and/or .flac files that may have originated from lossy codec sources:

http://losslessaudiochecker.com/
http://www.maurits.vdschee.nl/fakeflac/
http://tausoft.org/wiki/True_Audio_Checker_Algorithm

For the best audio quality, *do not accept MP3 audio from record companies or download services*. Even the best MP3 encoding is not audibly transparent with all program material, and there are a number of bad MP3 encoders in common use that produce even worse results.

MP4/AAC audio may be used on a case-by-case basis if nothing better is available. However, use of any source that been run through lossy compression will create potential problems with interaction between the source codec and transmission codec for media such as HD Radio® and streaming. This is discussed further in *Using Data Compression for Contribution* on page 22.

As of this writing, Amazon and Google Music sell downloads that use MP3 compression, while iTunes uses AAC compression. iTunes is thus preferred if an uncompressed or losslessly-compressed source (like .flac) is unavailable. (See *Lossless Compression* on page 12.)

If you use sources with lossy encoding like AAC, we recommend using a decoder that uses floating-point arithmetic or some other mechanism that prevents codec-induced overshoots from clipping the audio. Fixed-point decoding is acceptable if the decoder's designer has built adequate headroom into the decoder.

Sony/BMG Rootkit Malware

A scandal erupted in 2005 regarding Sony BMG's implementation of deceptive, illegal, and harmful copy protection measures on about 22 million CDs, representing 52 titles. When inserted into a computer, the CDs installed one of two pieces of software which provided a form of digital rights management (DRM) by modifying the operating system to interfere with CD copying. Neither program could easily be uninstalled, and they created vulnerabilities that were exploited by unrelated malware.[33]

A list of affected titles is available online[34]. It is unsafe to insert these CDs in a computer optical disk drive. Upon request, Sony/BMG will replace these titles with

---

[33] https://en.wikipedia.org/wiki/Sony_BMG_copy_protection_rootkit_scandal

[34] https://web.archive.org/web/20071012024250/http://cp.sonybmg.com/xcp/english/titles.html

"clean" re-issued CDs. Alternatively, they can be ripped via an analog transfer from a stand-alone CD player.

Mono Compatibility

Another pitfall in CD reissues is mono compatibility. This is important because most clock radios, smart speakers, and mobile phone speakers are mono and every FM car radio implements blend-to-mono as a function of signal strength and/or multipath. Each source that is transferred should be checked by ear to ensure that all audio channels (whether surround or stereo) sum to mono without artifacts. Sources that sound fine in their native formats may suffer from high frequency loss or "flanging" caused by uncorrected relative time delays between the audio channels. This can be caused by imperfections in analog recording, but also by "bit-slip" in buggy digital chains or in some old A/D converters (like the Sony PCM-F1) that multiplexed a single A/D converter between the left and right channels (see *bit-slip* on page 34). Some computer audio editing software, such as Adobe Audition, contains restoration tools like Automatic Phase Correction. With careful adjustment, possibly even in manual mode, good results are achievable.

Additionally, several Optimod processors, including Optimod-FM 8600, 8700i, and Optimod-PCn 1600, provide robust, automatic, on-line phase correction. The Optimod correction technology is multidimensional and is able to correct multiple problems simultaneously, such as phase cancellations caused by stereo micing of an instrument in the studio, cascaded with phase cancellation caused by gap skew in an analog tape recorder head.

Some mono CDs sourced from full-track mono masters have been transferred using a stereo playback tape head. The slight delay and azimuth differences between the two gaps can cause high frequency loss when summed to mono, which can be completely eliminated by choosing the best channel from the CD and using it as a source for both channels of the file in your playout system. However, if you can correct the delay errors by using the software or equipment mentioned above, you will gain 3 dB of signal-to-noise ratio by summing the channels after phase correction is applied.

Another issue is "electronically reprocessed for stereo" releases from the early days of stereo. These mostly had poor mono compatibility[35]. The worst example is processing that applied a time delay to the material in one channel with respect to the other. When summed to mono, such material exhibits extreme comb filtering. The

---

[35] An exception is material reprocessed using an Orban stereo synthesizer, which provides perfect mono compatibility and maintains the same subjective balance of mix elements as the mono source. This process is available in Optimod-PCn 1600. See Robert Orban, "A Rational Technique for Synthesizing Pseudo-Stereo from Monophonic Sources," JAES Volume 18 Issue 2 pp. 157-164; April 1970.

phase corrector in in the Optimod products mentioned above can correct this issue automatically.

Acoustic Summation vs. Electrical Summation: When considering mono compatibility, one subtlety to be aware of is the difference between acoustic and electrical summation of program elements in stereo or multichannel mixes. In most listening rooms, loudness of a given element in the mix is best represented as the sum of the power produced by that element in each loudspeaker because room acoustics tend to randomize the relative phase between the elements at the listening position. For example, when heard in stereo, an element present equally in the left and right channels of a stereo recording will sound about 3 dB louder than either channel by itself.

On the other hand, when the stereo channels are electrically summed to mono, elements add arithmetically, so in our example, the element will be 6 dB higher in the mono mix and the relative balance of the elements in the mono sum will not be the same as the stereo.

If you prefer to transmit a mono mix of a given track (for authenticity) but only have a stereo source, a partial solution to this dilemma is to pass the stereo source through a 90 degree phase difference network (mathematically termed a "Hilbert transformer") before it is summed to mono. This is best done by using a special FIR (finite-impulse-response) digital filter, where the left channel is delayed by the filter's group delay and the right channel is passed through the filter. Compared to using a pair of "allpass" filters (each with non-constant group delay) for the left and right channels, this this technique minimizes phase distortion, yet the filter's impulse response will still smear transient elements in the mix. If you use this technique it is important to listen carefully to decide if preserving the subjective balances in the stereo mix is more important than preserving the integrity of transients.

Early, ad-hoc "stereo" mixes of hit music from 1958 (when the stereo disk was first commercialized) to around 1966 were often not true stereo (in the sense of creating a continuous panorama between the left and right loudspeakers), but instead put elements only in the left or right channels—for example, vocals in the left and instruments in the right. This was the result of using three or four-track master recorders that did not have enough tracks to properly separate elements. Ironically, these recordings have the best mono compatibility in the sense that there is no change in balance between the elements when listening in stereo or mono. Some of these recordings were later mixed in true stereo by using digital source separation techniques to separate elements that were originally mixed on a given track of the multitrack master. For example, most of the Beatles' and many of the Beach Boys' recordings have remixed in this way.

Haeco CSG: Historically, the Haeco CSG ("Compatible Stereo Generator")[36] was created to process stereo recordings so that they maintained the same balance when

[36] https://en.wikipedia.org/wiki/Haeco-CSG

mixed to mono. It did this by introducing a relative phase shift between the channels (typically 90 degrees) and was in commercial use from about 1968 to 1970. Unfortunately, this processing not only blurs transients but also blurs the stereo image. It is wise to remove this processing in the production studio before a CGS-processed track is transferred to the playout system. This can be done by applying a 90 degree phase difference filter (so that all frequencies are 180 degrees out of phase) and then flipping the polarity of one channel to correct the 180 degree phase shift. Orban Optimod-PCn 1600 software for Windows includes a phase corrector that can accomplish this automatically, regardless of the amount of phase shift in the original CSG mastering. Some restoration software (like iZotope RX7 and Adobe Audition) has a phase shift module that allows manual correction.

Stereo Enhancement

In contemporary broadcast/netcast audio processing, high value is placed on the loudness and impact of a station compared to its competition. Orban originally developed the analog 222A Stereo Spatial Enhancer to augment a station's spatial image, achieving a more dramatic and more listenable sound. The stereo image becomes magnified and intensified; listeners also perceive greater loudness, brightness, clarity, dynamics and depth.

The 222A's technology detects and enhances the attack transients present in all stereo program material while not processing other portions. Because the ear relies primarily on attack transients to determine the location of a sound source in the stereo image, this technique increases the apparent width of the stereo soundstage. Because only attack transients are affected, the average L–R energy is not significantly increased, so the technique does not exacerbate multipath distortion.

While the 222A is no longer manufactured, several of Orban's digital Optimods now incorporate both the 222A algorithm and a delay-based algorithm in DSP.

Pre-Processing Files in Playout Systems

Applying full transmission-style audio processing (particularly multiband compression and peak limiting) to each file in a playout system is not recommended because unlike on-line processing of the final program, such processing cannot take into account the transitions between the different elements, particularly crossfades. Some material needs remastering (see *Restoration Software* below), but we recommend that for elements not requiring remastering, the only processing should be static loudness normalization to a BS.1770 Integrated target loudness, where a fixed gain adjustment (with proper dithering) is applied to the entire file to adjust the loudness to the target. This makes it easier for the online transmission processor to work smoothly and consistently.

A significant problem with normalizing all items in a playout system to a fixed BS.1770 Integrated loudness *without applying online transmission processing later* is that different types of material (particularly speech vs. music) require different targets to maximize listener comfort. The AES produced Recommendation TD1004.1.15-

10 to address this. See *Loudness Balance between Speech and Music* starting on page 38. Optimod online transmission processing uses several techniques, including an automatic speech/non-speech detector, to deal with this "genre" issue smoothly and automatically.

–23 LUFS is a good choice for the target loudness of files in playout systems, allowing ample peak headroom for almost any program material. (See *Measuring and Controlling Loudness* starting on page 35.) The only peak limiting that should be applied to program audio is peak limiting just before the transmitter or streaming encoder, and this should be done in an online transmission processor like an Optimod.

An exception to the –23 LUFS recommendation is classical music where the BS.1770 Integrated loudness of the source is below –23 LUFS. In this case, there are probably some peaks in the source that come close to 0 dBFS, so we suggest not increasing the loudness of such material.

Restoration Software

While vinyl records are particularly challenging, a significant amount of legacy music was recorded on analog equipment having an audible noise floor, like magnetic tape without complementary noise reduction systems (such as Dolby® or dbx®). Additionally, older recordings may have audible noise picked up in the recording studio, sum as hum from guitar amplifiers.

With the advent of "plug-in" signal processing architectures for both the PC and Mac platforms, DSP-based signal processing systems have become available at reasonable cost to remove tape hiss, hum, ticks, scratches, and noise from analog-sourced material, including vinyl disks. In a publication like this, designed for reasonably long shelf life, we can make few specific recommendations because the performance of the individual plug-ins is likely to improve quickly. These plug-ins typically cost a few hundred dollars, making them affordable to any production facility.

In addition to impulse noise reduction, such suites usually include an FFT-based dynamic noise reduction system to reduce low-level crackle, hiss, and rumble. These noise reduction systems typically use anywhere from 512 to 2048 frequency bands, enabling them to distinguish between noise and program material in a fine-grained manner and to subtract the estimated noise from the noisy signal. Most of the systems require the user to provide a "noise print" of typical noise (taken from a part of the groove with no program modulation), although the most advanced algorithms also provide a way to automatically estimate the noise print and to dynamically update it throughout the program being treated. These automatic systems are particularly valuable for vinyl noise reduction, where, unlike analog tape, the noise floor is unlikely to be statistically stationary.

Be aware that not all noise reduction systems or algorithms perform equally well. Some are dramatically more effective than others and some compromise audio quality more than others. Unsophisticated algorithms can suppress transients, distort brass and string sounds, and sound bubbly or flangy like MP3. In noise reduction algorithms the latter artifact is commonly termed “musical noise,” where the noise floor after processing emphasizes certain narrowband frequencies, which are perceived as having a musical pitch and thus sound unnatural. There are various techniques for ameliorating this[37] but not all software uses them, so choose your restoration software carefully.

Examples of affordable native restoration suites that we can recommend at this writing include iZotope Rx7[38] and Diamond Cut DC10[39]. At the high end, the line of hardware-based processors made by CEDAR®[40] in England has established itself as being the quality reference for this kind of processing. The CEDAR line is, however, *very* expensive by comparison to the plug-ins described above.

Other high-end products include the Sonic Solutions No-Noise® system (available as part of the Sonic Solutions workstations for mastering applications) and the TC Restoration Suite[41] for the Powercore Platform.

---

[37] For example, see: A. Lukin and J. Todd, ”Suppression of Musical Noise Artifacts in Audio Noise Reduction by Adaptive 2D Filtering,” Audio Engineering Society 123rd Convention, New York, NY, Oct. 2007

[38] http://www.izotope.com

[39] http://www.diamondcut.com

[40] http://www.cedar-audio.com

[41] http://www.tcelectronic.com

Editing in the Spectral Domain

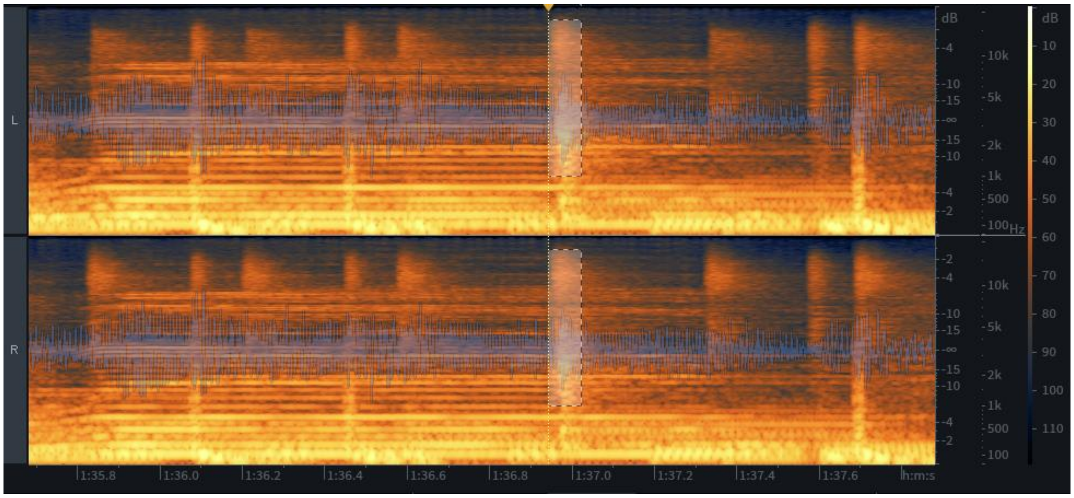

Figure 3: Editing the Spectral Domain

RX7 allows you to edit within a spectrogram view of the file and to superimpose the spectrogram on the time-domain waveform. A spectrogram displays the amount of energy in various frequency bands within the program material and how this energy varies over time. The amount of energy is indicated by the displayed local brightness.

Spectral editing facilitates manual removal of clicks, pops, and other noises, both impulsive and pitched, that are missed by automatic de-clicking algorithms. Unlike time-domain editors, which work on all frequencies simultaneously, spectral editing allows you to select only specified frequencies over specified periods of time. The screenshot in Figure 3 on page 73 shows a two-second chunk of a music file, with a dashed-line marquee delimiting the borders of the section of the spectrogram chosen for editing, which contains a snare drum hit. (The time domain waveform is shown faintly in blue behind the spectrogram; you can control the visual mix of the two as desired.) You could make the snare drum louder or quieter, or even remove it entirely via a spectral interpolation process that replaces the energy within the selection by energy interpolated from the selection's boundaries. Spectral editing can perform some remarkable tricks: for example, removing the sound of a smoke alarm or a cough from a live recording.

De-Clipping

De-clipping is a controversial technique that attempts to reconstruct the original audio waveform from a clipped or aggressively peak-limited version of it. There are many different de-clipping algorithms commercially available including tools in Adobe Audition, Steinberg WaveLab, iZotope RX, and Diamond Cut DC10. De-clipping has attracted interest because a distressing amount of contemporary source material has been "hypercompressed" in mastering or production, and record labels have then provided this material to broadcasters to use on-air, despite the fact that transmission audio processors often exaggerate the resulting audible distortion.

Information is 100% lost in flat-topped areas and cannot be recovered: A flat-topped waveform is a mathematical "singularity." Hence, de-clippers must make educated guesses about what's missing based on interpolation from material surrounding the clipped samples. To do this, the interpolation must use a model of the clipping process. However, many waveforms that look they have been hard-clipped have, in fact, been peak-limited by more complex limiting processes with sidechains and memory, and each limiter manufacturer has a proprietary way of computing the sidechain. For competitive reasons, these are seldom made public. Even if the sidechain is public knowledge, if the compression ratio is infinite, it is still impossible to deduce what the limiter's input was. For example, the "MX" peak limiter in the Orban Optimod-PCn 1600 software produces dense peaks very close to the value specified by its output level control, yet it uses a psychoacoustic model (similar to that used by a codec) to suppress audible distortion. Any "de-clipping" applied to this waveform will increase distortion. See Figure 4 on page 74. We would go so far as to state that when you are processing for loudness, if the waveform does not have frequent peaks at the threshold of the peak limiter (like Figure 4), the peak limiter is not using the available peak headroom in the channel efficiently and is likely to have side-effects like audible gain pumping.

De-clippers can increase punch on transients by increasing peak levels by guessing what the missing waveform is. However, this is not the same as cancelling IM distortion. Distortion cancellation depends on having a precise, invertible model of the peak limiting process. This is usually impossible. In fact, because de-clipping is a non-linear process, it can make its own IM distortion that adds to any IM distortion present in the original source. If simple peak clipping was used on a given track, then de-clippers can help, but sometimes they make things worse. The better the original peak limiting algorithm, the more likely it is that de-clipping will add IM distortion, not cancel it. Therefore, *the safest place for a de-clipper is in the production studio* (not the air chain), so that human ears can determine if the de-clipper is helping or

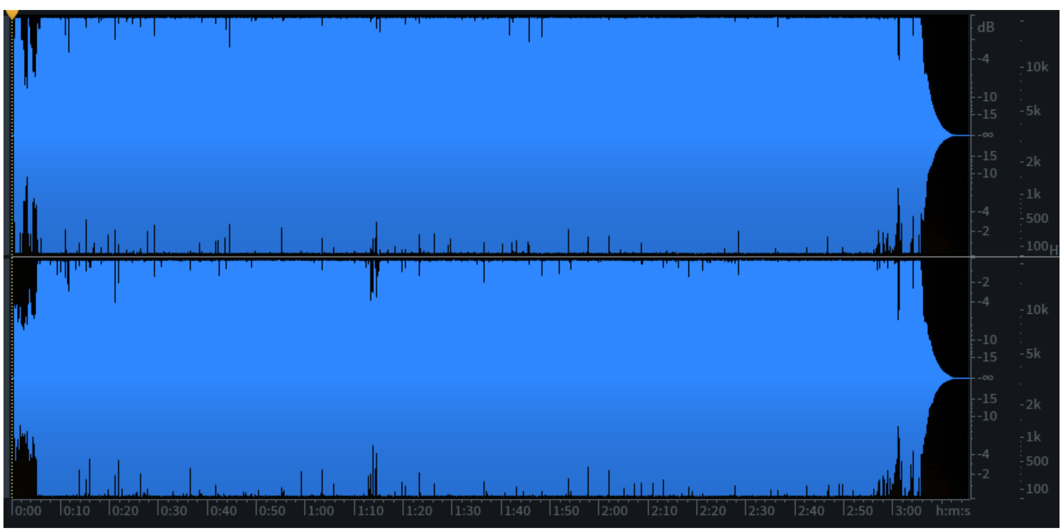

Figure 4: Program material mastered using Optimod-PCn 1600's MX Limiter

adding another layer of distortion. Moreover, in the broadcast processing chain, de-clipped waveforms force the on-air processor's peak limiter to work harder. So use de-clippers with care, and listen with your ears, not your eyes—waveforms like those in Figure 4 can be deceiving. Not all material that looks dense sounds bad.

Part 4: Equipment Following Optimod

Some of the equipment following Optimod in the transmission path can also affect quality. The STL, FM exciter, transmitter, and antenna can all have subtle, yet audible, effects.

STL

The availability of uncompressed digital STLs using RF signal paths has removed one of the major quality bottlenecks in the broadcast chain. These STLs use efficient modem-style modulation techniques to pass digitized signals with bit-for-bit accuracy. If the user uses their digital inputs and outputs and does not require them to do sample rate conversion (which can introduce overshoot if it a downward conversion that filters out signal energy), they are essentially transparent.

Uncompressed digital STLs using terrestrial lines (like T1s in the United States) also provide transparent quality and are equally recommended.

Some older digital STL technology uses lossy compression. If the bitrate is sufficiently high, these can be quite audibly transparent. However, all such STLs introduce overshoot and are therefore unsuitable for passing processed audio that has been previously peak limited.

Analog microwave STLs provide far lower quality than either digital technology and are not recommended when high audio quality is desired. They are sometimes appropriate for AM, because receiver limitations will tend to mask quality limitations in the STL.

Recently, the industry has informally implemented a digitized FM composite baseband connection using the left channel of a 192 kHz AES3 link. Because traditional analog composite connections are simple and robust, the main advantages of digitizing the composite are (1) increasing the resistance of the link to noise and EMI, and (2) allowing the entire signal path from studio to transmitter to remain in the digital domain.

Orban has extended this implementation to 384 kHz sampling by using the right channel of the link to pass even samples of 384 kHz while the left channel passes odd samples. This allows the full FM baseband (0-99 kHz) to be accommodated on the link. If the bandwidth of the original baseband signal is limited to 96 kHz, this signal is 100% backward-compatible with the implementation that uses the left channel only at 192 kHz.

FM Exciter

Exciter technology has improved greatly since FM's early years. The most important improvement has been the introduction of digitally synthesized exciters from several manufacturers. This technology uses no AFC loop and can have frequency response to DC if desired. It therefore has no problems with bounce or tilt to cause overshoot.

In conventional analog exciter technology, the major improvements have been lowered non-linear distortion in the modulated oscillator, and higher-performance Automatic Frequency Control (AFC) loops with better transient response and lower low-frequency distortion.

At this writing, the state-of-the-art in analog modulated oscillator distortion is approximately 0.02% THD at ±75 kHz deviation. (Distortion in digital exciters is typically 10 times lower than this.) In our opinion, if the THD of your exciter is less than 0.1%, it is probably adequate. If it is poorer than this (as many of the older technology exciters are), replacing your exciter will audibly improve sonic clarity and will also improve the performance of any subcarriers.

Even if the distortion of your modulated oscillator is sufficient, the performance of the AFC loop may not be. A high-performance exciter must have a dual time-constant AFC loop to achieve satisfactory low-frequency performance. If the AFC uses a compromise single time-constant, stereo separation and distortion will be compromised at low frequencies. Further, the exciter will probably not accurately reproduce the shape of the carefully peak- controlled Optimod-FM output, introducing spurious peaks and reducing achievable loudness.

Even dual time-constant AFC loops may have problems. If the loop exhibits a peak in its frequency response at subsonic frequencies, it is likely to "bounce" and cause loss of peak control. (Composite STLs can have similar problems.)[42]

Digital exciters have none of these problems. However, a *properly designed* analog exciter can have good enough performance to limit overshoot due to tilt and bounce to less than 1% modulation. Therefore, either technology can provide excellent results.

---

[42] Co-author Greg Oqonowski, Orban's Vice President of New Product Development, originally brought this to the industry's attention. (www.indexcom.com). Ogonowski has developed modifications for several exciters and STLs that improve the transient response of their AFCs.

FM Transmitter

The transmitter must be transparent to the modulated RF. If its amplifiers are narrowband (< 500 kHz at the -3dB points), it can significantly truncate the Bessel sidebands produced by the FM modulation process, introducing distortion. For best results, –3dB bandwidth should be at least 1MHz.

Narrowband amplifiers can also introduce synchronous FM. This can cause audible problems quite similar to multipath distortion, and can particularly damage SCAs. Synchronous FM should be *at least* –35dB below carrier level, with –40dB or better preferred.[43]

If the transmitter's group delay is not constant with frequency, it can also introduce synchronous FM, even if the bandwidth is wide. Please note that the "Incidental FM" reading on most FM modulation monitors is heavily smoothed and de-emphasized, and cannot be used to measure synchronous FM accurately. At least one device has appeared to do this accurately (Radio Design Labs' Amplitude Component Monitor Model ACM-1).

FM Antenna

Problems with antenna bandwidth and group delay can also cause synchronous FM, as can excessive VSWR, which causes reflections to occur between transmitter and antenna.

Perhaps the most severe antenna-induced problems relate to coverage pattern. Proper choice of the antenna and its correct installation can dramatically affect the amount of multipath distortion experienced by the listener. Multipath-induced degradations are far more severe than *any* of the other quality-degrading factors discussed in this paper. Minimization of received multipath is the single most important thing that the broadcast engineer can do to ensure high quality at the receiver.

AM Transmitter

We live in the golden age of AM transmitters. After 75 years of development, we finally have AM transmitters (using digital modulation technology) that are audibly transparent, even at high power levels. Previously, even the best high-power AM transmitters had a sound of their own, and all audibly degraded the quality of their inputs.

---

[43] Geoff Mendenhall of Harris has written an excellent practically-oriented paper on minimizing synchronous FM: G. Mendenhall, "Techniques for Measuring Synchronous FM Noise in FM Transmitters," *Proc. 1987 Broadcast Engineering Conf., National Assoc. of Broadcasters, Las Vegas, NV,* pp. 43-52 (Available from NAB Member Services)

We recommend that any AM station that is serious about quality upgrade to such a transmitter. By comparison to any tube-type transmitter, not only is the quality audibly better on typical consumer receivers, but the transmitter will pay for itself with lower power bills.

AM Antenna

The benefits of a transmitter with a digital modulator will only be appreciated if it feeds an antenna with wideband, symmetrical impedance. A narrowband antenna not only audibly reduces the high frequency response heard at the receiver, but also can cause non-linear distortion in radios' envelope detectors if asymmetrical impedance has caused the upper and lower sidebands to become asymmetrical. Such antennas will not work for any of the AM IBOC systems proposed at this writing.

Correcting antennas with these problems is specialized work, usually requiring the services of a competent consulting engineer.

DAB/ HD Radio / Netcasting Encoders

Most often, netcasts and podcasts use lossy compression at bitrates below 64 kbps. At these bitrates, audio quality depends critically on the choice of audio codec. At this writing, the highest quality codec at bitrates of 24 to 64 kbps codec is xHE-AAC. Refer to *Data Compression* on page 12 for a detailed discussion of transmission codecs.

Be aware that not all codec implementations sound the same. Even though various implementations of a specific codec type may encode/decode audio in the same format, the various implementations may not all produce the same audio quality. These codecs are only as good as their software realizations and there are many poor implementations available, especially from the unlicensed, open-source software community.

DAB (formerly called Eureka147) used the MPEG 1 Layer 2 codec (commonly called "MP2"). This provides poor audio fidelity at 128 kbps and borders on unacceptable at rates of 96 kbps and below. Because of these problems, DAB has been upgraded to DAB+, which uses the HE-AACv2 codec to achieve much more RF spectral efficiency than DAB by putting three good-sounding stereo channels where one mediocre-sounding channel used to fit with DAB.

HD Radiouses a proprietary codec called HDC. iBiquity has not released details about it, although it is known to use some sort of Spectral Band Replication technology (see page 20). Its subjective performance is better than MP3 but not as good as HE-AACv1 or v2.

Audio Processing for Low Bitrate Digital Transmissions

It is important to minimize audible peak-limiter-induced distortion when one is driving a low bitrate codec because one does not want to waste precious bits encoding the distortion. Look-ahead limiting can achieve this goal; hard clipping cannot.

One can model any peak limiter as a multiplier that multiplies its input signal by a gain control signal. This is a form of amplitude modulation. Amplitude modulation produces sidebands around the "carrier" signal. In a peak limiter, each Fourier component of the input signal is a separate "carrier" and the peak limiting process produces modulation sidebands around each Fourier component.

Considered this way, a hard clipper has a wideband gain control signal and thus introduces sidebands that are far removed in frequency from their associated Fourier "carriers." Hence, the "carriers" have little ability to mask the resulting sidebands psychoacoustically. Conversely, a look-ahead limiter's gain control signal has a much lower bandwidth than that of a clipper and produces modulation sidebands that are less likely to be audible.

Simple wideband look-ahead limiting can still produce audible intermodulation distortion between heavy bass and midrange material. The look-ahead limiter algorithm in Optimods uses sophisticated techniques to reduce such IM distortion without compromising loudness capability.

Conventional AM, FM, or TV audio processors that employ pre-emphasis/de-emphasis and/or clipping peak limiters do not work well with perceptual audio coders such as AAC/HE-AACv1/v2. The pre-emphasis/de-emphasis limiting in these processors unnecessarily limits high frequency headroom. Further, their clipping limiters create high frequency components—distortion—that the perceptual audio coders would otherwise not encode.

In addition, several audio processing manufacturers offer pre-processing claimed to minimize codec artifacts at low bitrates. Orban's technology is called PreCode™. This manipulates several aspects of the audio to minimize artifacts caused by low bitrate codecs, ensuring consistent loudness and texture from one source to the next. PreCode includes special audio band detection algorithms that are energy and spectrum aware. This can improve codec performance on some codecs by reducing audio processing induced codec artifacts, even with program material that has been preprocessed by other processing than Optimod.

Summary

Maintaining a high level of broadcast/netcast audio quality is a very difficult task, requiring constant dedication and a continuing cooperation between the programming, engineering, and computer IT departments.

With the constantly increasing quality of home and mobile receivers and stereo gear, the broadcast audience more and more easily perceives the results of such dedication and cooperation. One suspects that in the future, FM, DAR, and netcasts will have to deliver ever-increasing quality to compete successfully with the many other program sources vying for audience attention, including CD's, DVD's, Blu-ray disks, digital audio, subscription television, direct satellite broadcast, DTV, streaming programming on the Internet, high resolution downloads and who knows how many others!

The human ear is astonishingly sensitive; perceptive people are often amazed when they detect rather subtle improvements to the broadcast audio chain while listening to an inexpensive car radio. Conversely, the FM broadcast/reception system can exaggerate flaws in audio quality. Audio processors (even Optimod) are especially prone to exaggerating such flaws.

In this discussion, we have tried to touch upon the basic issues and techniques underlying audio quality in radio operations, and to provide useful information for evaluating the cost-effectiveness of equipment or techniques that are proposed to improve audio quality. In particular, we concluded that today's high-quality IC opamps are ideally suited for use as amplification elements in broadcast, and that compromises in digital standards, computer sound cards, disk playback, and tape quality are all likely to be audible on the air. The all-digital signal path is probably the single most important quality improvement that a station can make, but the installing engineer must be aware of issues such as lossy compression (particularly when cascaded), word length, sample rate, headroom, jitter, and dither, and 0dBFS+-induced clipping.

Following the suggestions presented here will result in better broadcast/netcast audio quality—and that is a most important weapon in attracting and maintaining an audience that is routinely exposed to compact discs and other high-quality audio reproduction media.

Provide your audience with the best possible experience.

The future belongs to the quality-conscious.

Appendix: Analog Media

Authors' Note for the 2019 Edition:

This Appendix devotes considerable space to the vagaries of analog media—vinyl disk and analog tape—that are becoming less and less important in broadcast production. However, given that they exist and that archival material may be stored on such media, we have chosen to retain this material (with minor editing) in the current revision. Because these media are analog, they require far more tweaking and tender loving care than do the digital media discussed above. For this reason, the following sections are long and detailed.

Vinyl Disk

Some radio programming still comes from phonograph records—either directly, or through dubs. Not only are some club DJs mixing directly to broadcast/netcast from vinyl, but also some old recordings have not been re-released on CD. Even if they have been, a disturbing fact is that many recently remastered CDs sound far worse than the original vinyl releases. This section discusses how to accurately retrieve as much information as possible from the grooves of any record.

Vinyl disk is capable of very high-quality audio reproduction. Consumer equipment manufacturers have developed high-fidelity cartridges, pick-up arms, turntables, and phono preamps of the highest quality. Unfortunately, much of this equipment has insufficient mechanical ruggedness for the pounding that it would typically receive in day-to-day broadcast operations.

There are only two reasonably high-quality cartridge lines currently made in the USA that are generally accepted to be sufficiently durable for professional use: the Stanton and the Shure professional series. Although rugged and reliable, these cartridges do not have the clean, transparent operation of the best high-fidelity cartridges. This phono cartridge dilemma is the prime argument for transferring all vinyl disk material to digital media in the production studio, and broadcasting only from digital media. In this way, it is possible (with care) to use state-of-the-art cartridges, arms, and turntables in the dubbing process, which should not require the mechanical ruggedness needed for broadcast equipment.

Good, high quality turntables and tonearms have become a bit scarce. However, the Technics SP-10 and its associated base (SH-10B3) and tonearm (EPA-B500/EPA-A250/EPA-A500) are very good choices for mastering vinyl to digital. This reduces the problem of record wear as well.

Production facilities specializing in high-quality transfer of vinyl to digital media should consider supplementing their conventional turntable with an ELP Laser Turntable[44]. Instead of playing disks mechanically, this pricey device plays vinyl without

[44] http://www.elpj.com/

mechanical contact to the disk, using laser beams instead. The authors have thoroughly evaluated the ELP and we can recommend it as delivering higher audio quality than any other vinyl playback device known to us.

Despite its “close to the master tape” sound quality, the laser turntable has several drawbacks. It is very sensitive to dust and imperfections in the grooves of a disk, so a wet vacuum cleaning (using a machine like a Loricraft, Nitty Gritty, or VPI) prior to playback is unconditionally required. (Of course, any archival transfer of vinyl should start with such a cleaning regardless of the playback technology employed.) The laser turntable will not play certain out-of-standard records, such as records where the cut starts on the outside raised bead, and its trackability is average — it will not track extremely high groove velocities that a state of the art cartridge can readily handle. Finally, it will not track non-black vinyl, such as picture disks. For these reasons, it cannot entirely supplant mechanical playback. However, it will correctly play a great majority of disks, and it can work wonders by ignoring surface damage (such as shallow scratches) that conventional playback will reveal.

Another important accessory for the specialist vinyl archiver (particularly when using the Laser Turntable) is a digital de-clicker and noise reduction system. (See *Restoration Software* on page 71.)

The following should be carefully considered when choosing and installing conventional vinyl disk playback equipment:

1. **Align the cartridge with great care.**

   When viewed from the front, the stylus must be absolutely perpendicular to the disc, to sustain a good separation. To prevent a fixed tracking error, the cartridge must be parallel to the headshell. Overhang should be set as accurately as possible ±1/16-inch (0.16 cm) and the vertical tracking angle should be set at 20° (by adjusting arm height). A valuable tool for precision alignment is the Protractor NG, available from Dr. Feickert Analogue
   http://www.feickert.de/engl/protractor.html

2. **Adjust the tracking force correctly.**

   Usually, better sound results from tracking close to the maximum force recommended by the cartridge manufacturer. If the cartridge has a built-in brush, do not forget to compensate for it by adding more tracking force according to the manufacturer’s recommendations. Note that brushes usually make it impossible to “back-cue,” although this should not be done when transferring to digital anyway.

3. **Adjust the anti-skating force correctly.**

   The accuracy of the anti-skating force calibration on many pick-up arms is questionable. The best way to adjust anti-skating force is to obtain a test record with an extremely high-level lateral cut (some IM test records are suitable). Connect the left channel output of the turntable preamp to the horizontal input of an oscilloscope and the fight channel preamp output to the vertical input. Operate the scope in the X/Y mode, such that a straight line at a 45-degree angle is visible. If the cartridge mistracks asymmetrically (indicating incorrect anti-skating compensation), then the scope trace will be "bent" at its ends. If this happens, adjust the anti-skating until the trace is a straight line (indicating symmetrical clipping).

It is important to note that in live-disk operations, use of anti-skating compensation may increase the chance of the phono arm sticking in damaged grooves instead of jumping over the bad spots. Increasing tracking force by approximately 15% has the same effect on distortion as applying anti-skating compensation. This alternative is recommended in live-disk operations.

4. Use a modern, direct-drive turntable.

None of the older types of professional broadcast turntables have low enough rumble to be inaudible for broadcast or netcast. These old puck-, belt-, or gear-driven turntables might as well be thrown away! Multiband audio processing can exaggerate rumble to extremely offensive levels.

Excellent consumer-style belt-drive turntables exist but they are not designed for slip-cueing, so they are only appropriate in the production studio.

5. Mount the turntable properly.

Proper turntable mounting is crucial—an improperly mounted turntable can pick up footsteps or other building vibrations, as well as acoustic feedback from monitor speakers (which will cause muddiness and severe loss of definition). The turntable is best mounted on a vibration isolator placed on a non-resonant pedestal anchored as solidly as possible to the building (or, preferably, to a concrete slab). The turntable bases supplied by the turntable manufacturer are highly recommended.

6. Use a properly adjusted, high-quality phono preamp.

Until recently, most professional phono preamps were seriously deficient compared to the best "high-end" consumer preamps. Fortunately, this situation has changed, and a small number of high-quality professional preamps are now available (mostly from small domestic manufacturers). A good preamp is characterized by extremely accurate RIAA equalization, high input overload point (better than 100mV at 1 kHz), low noise (optimized for the reactive source impedance of a real cartridge), low distortion (particularly CCIF difference-frequency IM), load resistance and capacitance that can be adjusted for a given cartridge and cable capacitance, and effective RFI suppression. The Firestone Audio Korora

RIAA Phono Stage Preamp and SUPPLIER Power Supply Unit, available from http://soundadditions.com, are highly recommended.

After the preamp has been chosen and installed, the entire vinyl disk playback system should be checked with a reliable test record for compliance with the RIAA equalization curve. (If you wish to equalize the station's air sound to produce a certain "sound signature," the phono preamp is *not* the place to do it.) Some of the better preamps have adjustable equalizers to compensate for frequency response irregularities in phono cartridges. Since critical listeners can detect deviations of 0.5dB, ultra-accurate equalization of the entire cartridge/preamp *system* is most worthwhile.

The load capacitance and resistance should be adjusted according to the cartridge manufacturer's recommendations, taking into account the capacitance of cables. If a separate equalizer control is not available, load capacitance and resistance may be trimmed to obtain the flattest frequency response. Failure to do this can result in frequency response errors as great as 10dB in the 10–15 kHz region! This is very often the reason many phono cartridge evaluations often produce colored results.

The final step in adjusting the preamp is to accurately set the channel balance with a test record, and to set gain such that output clipping is avoided on any record. If you need to operate the preamp close to its maximum output level due to the system gain structure, then observe the output of the preamp with an oscilloscope, and play a loud passage. Set the gain so that at least 6dB peak headroom is left between the loudest part of the record and peak-clipping in the preamp.

Choose a high quality A/D converter. Many computer onboard or internal sound cards produce excessive noise. Disable any soft clipping or limiting options in the A/D.

7. **Routinely and regularly replace styli.**

One of the most significant causes of distorted sound from vinyl disk reproduction is a worn phono stylus. Styli deteriorate sonically before any visible degradation can be detected even under a microscope, because the cause of the degradation is usually deterioration of the mechanical damping and centering system in the stylus (or actual bending of the stylus shank), rather than diamond wear. This deterioration is primarily caused by back-cueing, although rough handling will always make a stylus die before its time.

Styli used in 24-hour service should be changed every two weeks as a matter of course—whatever the expense! DJs and the engineering staff should listen constantly for audible deterioration of audio quality, and should be particularly sensitive to distortion caused by a defective stylus. *Immediately* replace a stylus when problems are detected. One engineer we know destroys old styli as soon as he replaces them so that he is not tempted to keep a stock of old, deteriorated, but usable-looking styli!

It is important to maintain a stock of new spare styli for emergencies, as well as for routine periodic replacement. There is no better example of false economy than waiting until styli fail before ordering new ones, or hanging onto worn-out styli until they literally collapse! Note also that smog- and smoke-laden air may seriously contaminate and damage shank mounting and damping material. Some care should be used to seal your stock of new styli to prevent such damage.

Analog Tape

Despite its undeniable convenience, the tape cartridge (even at the current state of the art) is inferior to reel-to-reel in almost every performance aspect. Performance differences between cart and reel are readily measured, and include differences in frequency response, noise, high-frequency headroom, wow and flutter, and particularly azimuth and interchannel phasing stability. While tape cartridges are long obsolete for on-air operations, they are of interest if they hold archival material that is otherwise unavailable and that needs to be transferred to a playout system.

Cassettes were occasionally promoted as a serious broadcast program source. Cassettes' low speed, tiny track width, sensitivity to dirt and tape defects, and *substantial* high-frequency headroom limitations make such proposals totally impractical where consistent quality is demanded. Cassettes are mostly of interest today if they contain archival material unavailable of higher-fidelity media. It is generally accepted that the best quality cassette player ever made was the Nakamichi Dragon, which has been out of production since 1993 but sometimes available on the used market. A serious archivist would be well-advised to acquire one and if necessary, to have it restored.

Sum and Difference Recording:

Because it is vital in stereo FM broadcast to maintain mono compatibility, sum and difference recording is preferred in either reel or cart operations. This means that the mono sum signal (L+R) is recorded on one track, and the stereo difference signal (L–R) is recorded on the other track. A matrix circuit restores L and R upon playback. In this system, interchannel phase errors cause frequency-dependent stereo-field localization errors rather than deterioration of the frequency response of the mono sum.

Because this technique tends to degrade signal-to-noise (L+R usually dominates, forcing the L–R track to be under-recorded, thereby losing up to 6dB of signal to-noise ratio), it is important to use a compander-type noise reduction system if sum-and-difference operation is employed.

Electronic Phase Correction

Because interchannel phase errors are endemic on analog tape, it is wise to maintain a transfer machine in which the reproduce head azimuth adjustment is readily available for tweaking by ear. This is particularly effective if the technician listens to the sum of the channels and minimizes audible high frequency loss.

Orban makes several products that include a multidimensional phase corrector that will eliminate phase errors on cartridge playback. For the production studio, Optimod-PCn 1600 processing software is usually the most appropriate.

Cheap Tape:

Cheap tape, whether reel or cart, is a temptation to be avoided. Cheap tape may suffer from any (or all) of the following problems:

- Sloppy slitting, causing the tape to weave across the heads or (if too wide) to slowly cut away your tape guides.
- Poor signal-to-noise ratio.
- Poor high-frequency response and/or high-frequency headroom.
- Inconsistency in sensitivity, bias requirements, or record equalization requirements from reel to reel (or even within a reel).
- Splices within a reel.
- Oxide shedding, causing severe tape machine cleaning and maintenance problems.
- Squealing due to inadequate lubrication.

High-end, name-brand tape is a good investment. It provides high initial quality, and guarantees that recordings will be resistant to wear and deterioration as they are played. Whatever your choice of tape, you should standardize on a single brand and type to assure consistency and to minimize tape machine alignment problems. Some of the most highly regarded tapes in 1990 use included Agfa PEM468, Ampex 406, Ampex 456, BASF SPR-50 LHL, EMI 861, Fuji type FB, Maxell UD-XL, TDK GX, Scotch (3M) 206, Scotch 250, Scotch 226, and Sony SLH1 1. In 2014, none of these tapes are being manufactured. Before considering use of old stock of these tapes for new recordings, be aware that several suffered from "sticky-shed syndrome" (see page 92) and may have deteriorated severely since they were manufactured. It is safer to use newly manufactured tape for new recordings.

In late 2018, there were only a few remaining manufacturers of analog audio tape. Recording the Masters [45] makes open reel and cassette tape from the original AGFA

[45] https://www.recordingthemasters.com/

and BASF specifications. ATR Magnetics of York, PA, longtime service and modification shop for multitrack and master recorders, manufactures open-reel and cassette tapes. Jai Electronic Industries in India makes audio tape in 6.35 mm (1/4") and 12.7 mm (1/2") width, and perforated 16 mm and 35 mm audio tape for the film industry.[46]

Tape Speed:

If all aspects of the disk-to-tape transfer receive proper care, then the difference in quality between 15ips (38cm/sec) and 7.5ips (19cm/sec) recording is easily audible. 15ips has far superior high-frequency headroom. The effects of drop-outs and tape irregularity are also reduced, and the effects of interchannel phase shifts are halved. However, a playback machine can deteriorate (due to oxide build-up on the heads or incorrect azimuth) far more severely at 15ips than at 7.5ips before an audible change occurs in audio quality.

Noise Reduction:

A compander-type (encode/decode) noise reduction system can be used to reduce tape hiss to an unobtrusive or even inaudible level. However, if transparency is desired, it is difficult to imagine a contemporary broadcast application where companded analog tape would be preferred to linear PCM digital recording, which is reliably transparent when implemented correctly. In contemporary production, tape is usually used because it colors the sound in a way that artists and producers find attractive. Tape hiss, soft saturation, and modulation noise are part of that color. Because 7.5 ips introduces more high-frequency saturation than 15 ips, the production community has found both speeds to be useful for different effects.

However, compander technology is still of interest because many legacy recordings were recorded using a compander-type noise reduction system and correct playback requires access to the compander hardware or a digital model of it. We have evaluated and can enthusiastically recommend Dolby SR (Spectral Recording). Good results have been reported with Telcom C4 as well. dbx Type II noise reduction is also effective and has the advantages of economy and freedom from mistracking due to level mismatches between record and playback. Dolby A was the original multiband compander and many legacy recordings were recorded using the Dolby A system, which provides modest amounts of noise reduction but no audible noise pumping or breathing when properly aligned.

Remember that to achieve accurate Dolby tracking, record and playback levels must be matched within 2dB. Dolby noise (for SR operations), or the Dolby tone (for Dolby A operations) should always be recorded at the head of all reel-to-reel tapes, and level-matching should be checked frequently. There should be no problem with level-matching if tape machines are aligned every week, as level standardization is part of this procedure. If a different type of tape is put in service, recording machines must be aligned to the new tape *immediately,* before any recordings are made.

---

[46] http://en.wikipedia.org/wiki/Reel-to-reel_audio_tape_recording

In our opinion, all single-ended (dynamic noise filter) noise reduction systems can cause undesirable audible, program-dependent side-effects and cannot safely be used on-line. The best DSP-based systems can be very effective in the production studio (where they can be adjusted for each piece of program material), but even there they must be used carefully, with their operation constantly monitored by the station's "golden ears." Some possible applications include noise reduction of outside production work, and, when placed after the microphone preamp, reduction of ambient noise in the control room or production studio.

Tape Recorder Maintenance:

Regular maintenance of magnetic tape recorders is crucial to achieving consistently high-quality sound. Tape machine maintenance requires expertise and experience. The following points provide a basic guide to maintaining a tape recorder's performance.

1. **Clean heads and guides every four hours of operation.**

2. **Demagnetize heads as necessary.**

   Tradition has it that machines should be demagnetized every eight hours. In our experience, magnetization is usually not a problem in playback-only machines in fixed locations. A magnetometer with a ±5 gauss scale (available from R.B. Annis Co., Indianapolis, Indiana, USA) should be used to periodically check for permanent magnetization of heads and guides. You will find out how long it takes for *your* machines in *your* environment to pick up enough permanent magnetization to be harmful. You may well find that this never happens with playback machines. Recording machines should be watched much more carefully.

3. **Measure tape machine performance frequently.**

   Because tape machine performance usually deteriorates gradually, measure the performance of broadcast machines frequently with standard test tapes. Take whatever corrective action is necessary if the machine is not meeting specifications. As of 2014, test tapes are still available from Magnetic Reference Laboratory (MRL) (165 Wyandotte Dr, San Jose, CA 95123; www.mrltapes.com)

4. **Measure flutter.**

   Routine maintenance should include measurement of flutter, using a flutter meter and high-quality test tape. Deterioration in flutter performance is often an early warning of possible mechanical failure. Spectrum analysis of the flutter can usually locate the flutter to a single rotating component whose rate of rotation corresponds to the major peak in the filter spectrum. Deterioration in flutter performance can, at very least, indicate that adjustment of reel tension, capstan tension, reel alignment, or other mechanical parameter is required.

5. Measure frequency response and interchannel phase shifts.

These measurements, which should be done with a high-quality alignment tape, can be expedited by the use of special swept frequency or pink noise tapes available from some manufacturers (like MRL). The results provide an early indication of loss of correct head azimuth, or of headwear. (The swept tapes are used with an oscilloscope; the pink noise tapes with a third-octave real time analyzer.)

The head must be replaced or lapped if it becomes worn. Do not try to compensate by adjusting the playback equalizer. This will increase noise unacceptably, and will introduce frequency response irregularities because the equalizer cannot accurately compensate for the shape of the rolloff caused by a worn head.

6. Record and maintain alignment properly.

Alignment tapes wear out. With wear, the output at 15 kHz may be reduced by several dB. If you have many tape machines to maintain, it is usually more economical to make your own "secondary standard" alignment tapes, and use these for weekly maintenance, while reserving your standard alignment tape for reference use. (See below.) However, a secondary standard tape is not suitable for critical azimuth adjustments. These should be made using the methods described above, employing a test tape recorded with a full-track head. Even if you happen to have an old full-track mono machine, getting the azimuth *exactly* right is not practical—use a standard commercial alignment tape for azimuth adjustments.

The level accuracy of your secondary standard tape will deteriorate with use—check it frequently against your primary standard reference tape. Because ordinary wear does not affect the azimuth properties of the alignment tape, it should have a very long life if properly stored.

Store all test tapes:

- Tails out.
- Under controlled tension.
- In an environment with controlled temperature and humidity.
- With neither edge of the tape touching the sides of the reel (this can only be achieved if the tape is wound onto the storage reel at normal playback/record speeds, and *not* at fast-forward or rewind speed).

7. Check playback alignment.

A) Coarsely adjust each recorder's azimuth by peaking the level of the 15 kHz tone on the alignment tape.

Make sure that you have found the *major* peak. There will be several minor peaks many dB down, but you will not encounter these unless the head is totally out of adjustment.

B) While playing back the alignment tape, adjust the recorder's reproduce equalizers for flat high-frequency response, and for low-frequency response that corresponds to the fringing table supplied with the standard alignment tape.

Fringing is due to playing a tape that was recorded full-track on a half track or quarter-track head. The fringing effect appears below 500Hz, and will ordinarily result in an apparent bass boost of 2-3dB at 100Hz.

Fine azimuth adjustment cannot be done correctly if the playback equalizers are not set for identical frequency response, since non-identical frequency response will also result in non-identical phase response.

C) Fine-adjust the recorder's azimuth.

This adjustment is ideally made with a full-track mono pink noise tape and a real-time analyzer. If this instrumentation is available, sum the two channels together, connect the sum to the real-time analyzer, and adjust the azimuth for maximum high-frequency response.

If you do not have a full-track recorder and real-time analyzer, you could either observe the mono sum of a swept-frequency tape and maximize its high-frequency response, or align the master recorder by ear. Adjust for the crispest sound while listening to the mono sum of the announcer's voice on the standard alignment tape (the azimuth on the announcer's voice will be just as accurate as the rest of the tape).

If the traditional Lissajous pattern is used, use *several* frequencies, and adjust for minimum differential phase at *all* frequencies. Using just one frequency (15 kHz, for example) can give incorrect results.

8. Check record alignment, and adjust as necessary.

Set record head azimuth, bias, equalization, and calibrate meters according to the manufacturer's recommendations. We recommend that tape recorders be adjusted so that +4dBu (or your station's standard operating level) in and out corresponds to 0VU on the tape recorder's meters, to Dolby level, and to standard operating level. (This is ordinarily 250 nW/m for conventional tape and 315 nW/m for high output tape—refer to the tape manufacturer's specifications for recommended operating fluxivity.)

Current practice calls for adjusting bias with the "high frequency overbias" method (rather than with the prior standard "peak bias with 1.5-mil wavelength" method). To do this, record a 1.5-mil wavelength on tape (5 kHz at 7.5ips) and increase the bias until the maximum output is obtained from this tape. Then *further* increase the bias until the output has decreased by a fixed amount, usually 1.5 to 3dB (the correct amount of decrease is a function of both tape formulation and the width of the gap in the record head—consult the tape manufacturer's data sheet)

9. Follow the manufacturer's current recommendations

In addition to the steps listed above, most tape machines require periodic brake adjustments, reel holdback tension checks, and lubrication. With time, critical bearings will wear out in the motors and elsewhere (such failures are usually indicated by incorrect speed, increased flutter, and/or audible increases in the mechanical noise made by the tape recorder). Use only lubricants and parts specified by the manufacturer.

10. Keep the tape recorder and its environment clean.

Minimize the amount of dust, dirt, and even cigarette smoke that comes in contact with the precision mechanical parts. In addition to keeping dust away from the heads and guides, periodically clean the rest of the machine with a vacuum cleaner (in *suction* mode, please!), or with a soft, clean paintbrush. It helps to replace the filters in your ventilation system at least five times per year.

Recording Your Own Alignment Tapes

Recording a secondary standard alignment tape requires considerable care. We recommend you use the traditional series of discrete tones to make your secondary standard tapes.

A) Using a standard commercial alignment tape, very carefully align the playback section of the master recorder on which the homemade alignment tape will be recorded (see step 7 on page 89).

> *While aligning the master recorder,* write down the actual VU meter reading produced at each frequency on the spot-frequency standard alignment tape.

B) Subtract the compensation specified on the fringing table from the VU meter readings taken in step (A).

> Because you are recording in half-track stereo instead of full-track mono, you will use these compensated readings when you record your secondary standard tape.

C) Excite the record amplifier of the master recorder with pink noise, spot frequencies, or swept tones.

D) Adjust the azimuth of the master recorder's record head, by observing the mono sum from the playback head.

> Pink noise and a real-time analyzer are most effective for this.
>
> If the traditional Lissajous pattern is used, use *several* frequencies, and adjust for minimum differential phase at *all* frequencies.

E) Set the master recorder's VU meter to monitor playback.

F) Record your secondary standard alignment tape on the aligned master recorder.

Use an audio oscillator to generate the spot frequencies. *Immediately* after each frequency is switched in, adjust the master tape recorder's record gain control until the VU meter reading matches the *compensated* meter readings calculated in step (B).

Your homemade tape should have an error of only 0.5dB or so if you have followed these instructions carefully.

"Sticky Shed Syndrome"

Tape manufactured from the 1970s through the 1990s (particularly by AGFA, Ampex, and 3M) may suffer from so-called "sticky shed syndrome." When played, the tape sticks to the guides of the playback machine and severe oxide loss may occur.

The generally accepted cure is to bake the tape at 130° F (54° C) in a convection oven. One recommended device is the Snackmaster Pro model FD-50 made by American Harvest[47]. Don't use the oven in a household stove or a microwave oven. Baking time ranges from about 4 hours for ¼" tape to 8 hours for 2" tape, although it's not critical. You can't over-bake unless you leave the tape in for a day or so; if you under-bake and the tape is still gummy, you can bake it more. After you shut off the heat, leave the tape to cool down to room temperature before attempting to play it.

A baked tape should be playable for about a month, although this depends greatly on the ambient humidity. Although many tapes can be re-baked as necessary, this is not always true; baking has risks[48]. It is important to make a high-quality digital archive of the tape on its first pass through the playback machine after baking. This will minimize the probability that the tape will suffer catastrophic damage later on[49].

---

[47] (800 288-4545; www.americanharvest.com). Model FD-50 is no longer being manufactured. However, American Harvest still makes similar products, which we have not evaluated.

[48] Bill Holland, "Industry's Catalog at Risk – Archived Tapes could be Lost to Binder Problem," Billboard Magazine, June 5, 1999. (This article is not available on line unless you subscribe to Billboard's online service, so a local library may be the best way of getting it.)

[49] Useful discussions of sticky shed syndrome can be found at:
http://www.clir.org/pubs/reports/pub54/2what_wrong.html and
http://mixonline.com/ar/audio_sleep_egyptian/

Cartridge Tape Machine Maintenance:

The above comments on tape recorder maintenance apply to cart machines as well. However, cart machines have further requirements for proper care—largely because much of the tape guidance system is located *within the cartridge,* and so is quite sensitive to variations in the construction of the individual carts.

While few cart machines are still in use (almost entirely to retrieve archival material that is otherwise unavailable), some broadcasters have found that the heftier ones make good doorstops.

1. Clean pressure rollers and guides frequently.

Because lubricated tape leaves lubricant on the pressure rollers and tape guides, frequent cleaning is important in achieving the lowest wow and flutter and in preventing possible cartridge jams. Cleaning should be performed as often as experience proves necessary. Because of the nature of tape lubricant, it does *not* tend to deposit on head gaps, so head cleaning is rarely required.

2. Check head alignment frequently.

Even with the best maintenance, interchannel phase shifts in conventional cart machines will usually prove troublesome. In addition, different brands of cartridges will show significant differences in phase stability in a given brand of machine. Run tests on various brands of carts, and standardize on the one offering best phase stability.

3. Follow the manufacturer's maintenance and alignment instructions.

Because of the vast differences in design from manufacturer to manufacturer, it is difficult to provide advice that is more specific.

4. Consider upgrading the cart machine's electronics.

Many early (and some not-so-early) cart machines had completely inadequate electronics. The performance of these machines can be improved considerably by certain electronics modifications. Check the machine for the following:

A) record-amplifier headroom (be sure the amplifier can completely saturate the tape before it clips)

B) record amplifier noise and equalization (some record amplifiers can actually contribute enough noise to dominate the overall noise performance of the machine)

C) playback preamp noise and compliance with NAB/IEC equalization

D) power supply regulation, noise, and ripple

E) line amplifier headroom

F) record level meter alignment (to improve apparent signal-to-noise ratio at the expense of distortion, some meters are calibrated so that 0 corresponds to significantly more than 1% third-harmonic distortion!)

Probably the most common problem is inadequate record amplifier headroom. In many cases, it is possible to improve the situation by increasing the operating current in the final record-head driver transistor to a value close to its power dissipation limits. This is usually done by decreasing the value of emitter (and sometimes collector) resistors while observing the collector voltage to make sure that it stays at roughly half the power supply voltage under quiescent conditions, and adjusting the bias network as necessary if it does not.

About the Authors

Robert Orban

Robert Orban received the B.S.E.E. degree from Princeton University in 1967 and the M.S.E.E. degree from Stanford University in 1968.

In 1970, he founded Orban Associates, originally as a manufacturer of studio equipment. In 1975, Orban Associates introduced the original Optimod-FM 8000, which was the first in a long line of broadcast audio processors for AM, FM, TV, and digital broadcasting from the company. Although Orban Associates changed hands several times in the ensuing decades, Orban continues to work at the successor company as Vice President/Chief Engineer.

Starting at age six, he began studying piano and voice. A knack for improvisation and musical composition led him at age 15 to create a setting of Psalm 108 for mixed chorus and piano, which was performed publicly to considerable acclaim by the chorus at his high school in Butler, New Jersey. At Princeton, in addition to the studies that led to his EE degree, he took the first three years of the composition and music theory courses usually taken by music majors.

Orban has been involved in professional recording for many years. In Princeton, he supplemented his income by offering on-location recording services, recording many performances for both for the 17 kW commercially licensed campus radio station WPRB(FM), and for the University itself. He was closely associated with WPRB throughout his college years, hosting a weekly classical music show and serving at various times as Chief Engineer and Music Director. He also became a skilled top-40 board operator when this meant relying on live talent, cart machines and slip-cued vinyl. It was at WPRB that he designed and built his first audio processor and caught the "radio bug" that led to his career as a successful broadcast equipment designer.

Around 1970, he became associated with electronic music pioneers Paul Beaver and Bernie Krause and mixed several of their records for the Warner Bros. label. Later, he worked with a number of independent artists and labels as a writer, mixer, instrumentalist, and producer. Orban's compositions been heard on classical radio stations in New York and San Francisco, and his score for a short film, "Dead Pan," was heard on PBS television in Chicago. He was able to exploit his experience in pro audio when designing studio reverberators, stereo synthesizers, compressors, parametric equalizers, enhancers, and de-essers under both the Orban and dbx brand names.

Orban has been actively involved in NRSC committee AM improvement work. He has been widely published in both the trade and refereed press (including J. Audio Engineering Soc., Proc. Soc. Automotive Engineers, and J. SMPTE). He co-authored the chapter on "Transmission Audio Processing" in the NAB Engineering Handbook, 11th edition. He currently holds over 25 U.S. patents.

In 1973, he was elected a Fellow of the Audio Engineering Society. In 1993, he shared with Dolby Laboratories a Scientific and Engineering Award from the Academy of Motion Picture Arts and Sciences. In 1995, he received the NAB Radio Engineering Achievement Award. In 2002, he received the Innovator award from Radio Magazine. Today, he continues to actively research new DSP audio processing technology and to write, produce, and record music.

Greg Ogonowski

Having studied piano as a child, Greg began to appreciate the sound quality and musicality of various sources. It was his ear candy, and his infatuation with commercial Top 40 radio would eventually help shape and mold the way radio audio processing sounds today.

It was the summer of 1967. A Los Angeles based group called the Fifth Dimension had a hit called "Up, Up, and Away." Greg noticed it had a fresh, new sound, thanks to engineer Bones Howe, and Greg managed to catch it playing on CKLW Radio and WJR Radio at the exact same time. He switched between the stations numerous times, fascinated by how different the two sounded. He could not understand why WJR was allowed to sound so inferior to the BIG full sound of CKLW, whose audience share was growing daily. This seemingly small experience was about to change Greg's life and would ultimately affect the technical sound of radio broadcasting and Internet netcasting as we now know it.

It took several years of inspiration and hearing the hits from Motown and the LA Wrecking Crew before he understood what would be needed to craft a broadcast audio processing system that would make this music shine over the air by creating a big, consistent sound that lived up to or surpassed the standards set by CKLW. His first radio gig was at WWWW, aka "W4," in Detroit while he was still in high school. Here he learned the ins and outs of the commercial broadcast business, including the technical workflow.

In 1975, he founded Gregg Laboratories, a broadcast audio signal processing company, and he has had considerable experience designing commercial broadcast audio processing systems that many high profile broadcasters have used. He has extensively researched the characteristics of consumer radio receivers and, with Robert Orban, co-authored a technical paper proposing the standardization of pre-emphasis in AM broadcast. This paper was presented before the Society of Automotive Engineers and National Association of Broadcasters and was the precursor to the AM-NRSC standard.

He has also authored and co-authored many other technical papers on various topics relating to the audio and broadcast industry.

In 1984, he founded Modulation Index, a broadcast engineering consultancy, and has done studies on broadcast modulation measurement instrumentation and FM modulators, including STL's and exciters. As a result of these studies, he developed modifications for popular monitors, STL'S, and exciters to improve their dynamic transient accuracy and competitiveness, and he presented a technical paper before the National Association of Broadcasters regarding these findings. He developed audio signal processing algorithms that were later included in all of the current generation of Optimod audio processors.

As technical director for KTNQ/KLVE, Heftel Broadcasting, Los Angeles, from 1985 to 1991, he relocated studio facilities and constructed a new efficient alternative use AM transmission facility. As technical director for KBIG/KLAC, Los Angeles from 1998 to 2000, he installed a new computer network and digital audio delivery system throughout the business and technical facility. KBIG/KLAC was one of the first radio stations to stream audio on the Internet with high fidelity sound, all from internal encoders and servers.

After Ogonowski joined Orban in 2000 as VP Product Development, he led the team that created Optimod-PC, a PCI sound card with onboard audio processing for any digital audio or streaming application. He currently oversees the engineering department, where audio encoders, editors, and signal processors currently under development will enable Orban to continue in its tradition of high quality, high performance broadcast technology.

Determined to change the way Internet streaming audio is perceived and consumed, he was the architect of the first commercial high quality file and streaming audio encoder using standards-based MPEG-4 AAC/HE-AAC and MPEG Surround, Orban Opticodec-PC. He also created a high quality HE-AAC streaming player supporting standards-based protocols, the StreamS HiFi Radio App, which was the first Adobe Flash streaming audio player for Apple iPhone/iPad.

Index

B

C

D

E

F

G

P

Q

R

S

V

W

X

Y