QuickEnhance®

Speech Clarification Processor
Forensic AudioSuite™ and VST Plug-in

User's Manual

Digital Audio Corporation

ii

QuickEnhance®

Speech Clarification Processor
Forensic AudioSuite™ and VST Plug-in

User's Manual

Digital Audio Corporation
4018 Patriot Drive
One Park Center, Suite 300
Durham, NC 27703
Phone: 919 572 6767
Fax: 919 572 6786

Web: www.dacaudio.com
Email: sales@dacaudio.com

TABLE OF CONTENTS

1   QUICKENHANCE INTRODUCTION

1.1   OVERVIEW OF THE QUICKENHANCE SYSTEM

Thank you for purchasing the QuickEnhance Voice Enhancement System from Digital Audio Corporation. QuickEnhance provides easy-to-use filtering options for noise removal to help improve speech intelligibility and clarity in noisy audio. QuickEnhance is a forensic plug-in that has been specifically designed to work with AudioSuite and VST host audio editing systems.

QuickEnhance comes equipped with the following features:

- Auto-Normalizer
- 200 Hz High-pass Filter
- 1000-point Spectrum Analyzer (with maximum peak indicator)
- Hum Filter (adjustable base frequency with Auto-Tracking feature)
- Predictable Noise Reduction Filter (Deconvolver)
- Random Noise Reduction Filter (Broadband filter)
- 20-Band Graphic Equalizer
- Automatic Gain Control

2 SOFTWARE INSTALLATION

2.1 QUICKENHANCE SOFTWARE INSTALLATION

The QuickEnhance software requires that an host audio editing environment that supports either the AudioSuite or VST standard be installed. The computer on which the audio editing environment is installed must meet all the requirements of that audio editing environment for the QuickEnhance plug-in to work. Please consult the documentation of your audio editing environment for its requirements.

NOTE: Because of the processing requirements of the plug-in, the filters will benefit from a faster computer.

The software package consists of the QuickEnhance/AS and QuickEnhance/VST plug-in developed by DAC. To start the installation, insert the CD containing the QuickEnhance software into your PC. If the installation does not begin automatically run the file "Launcher.exe" located in the root folder of the CD. You can run this file by clicking on the "Start" menu, then on "Run", and typing "X:\Launcher.exe" (where "X" is the drive letter of your CD-ROM drive). You will be prompted to install either the QuickEnhance/AS or QuickEnhance/VST version. If you wish to install both, simply run the Launcher executable again after finishing the first installation. After your selection, follow the instructions on the screen to complete the installation. If you have any questions or problems with the installation, please contact DAC for further assistance.

The AudioSuite plug-in will be installed to the default Digidesign plug-in folder (normally, "C:\Program Files\Digidesign\DAE\Plug-Ins"). If this folder is not found, it will be created. Remember to <u>first</u> install your audio editing environment that supports Digidesign AudioSuite plug-ins before installing the QuickEnhance/AS.

The VST plug-in will be installed to the default Steinberg VST plug-in folder (normally, "C:\Program Files\Steinberg\VstPlugins"). If this folder is not found, it will be created.

3   QUICKENHANCE FUNCTIONALITY

QuickEnhance is an audio processing plug-in designed for use in environments that support the AudioSuite™ or VST plug-ins. QuickEnhance provides a set of speech clarification tools, primarily for use in forensic applications.

Plug-ins are launched from host applications in various ways depending on the host. To launch QuickEnhance from within Digidesign Pro Tools®, select the audio you would like to clarify and then choose QuickEnhance from the AudioSuite menu. To launch QuickEnhance from within Avid XPress® Pro, open the AudioSuite Window by selecting AudioSuite from the Tools menu. Place your audio in the timeline (or in the AudioSuite window, if you are running in Master Clip mode). Then select QuickEnhance from the plug-in drop-down selector.

For VST hosts, usually the plug-in can be found in the "Effects", "Tools" or "VST" menu. Please consult your audio editor manual to determine how to recognize and launch a VST plug-in.

A screenshot of the QuickEnhance/AS main window (as displayed in Avid XPress® Pro) is shown below. In the Help index, you will find links to pages that describe each of the filter control panels shown in the main window.
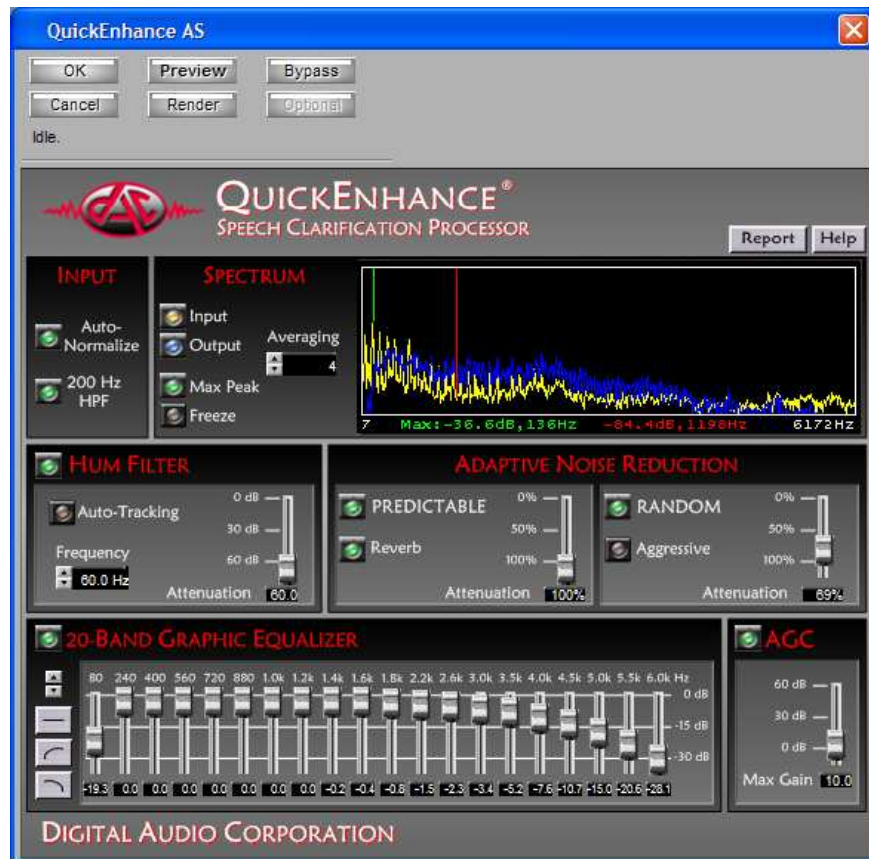
Figure 1: QuickEnhance/AS Main Window

3.1 MAIN PLUG-IN WINDOW

Figure 1 shows the plug-in as it would be displayed if used in Avid XPress® Pro. The buttons above and below the QuickEnhance Main Plug-in Window will appear differently depending upon your audio editing environment. Please consult the documentation on your specific audio editing environment for details of how to use these features.

The Main Window Plug-in Window provides access to all of the audio filters. The available filters are:

- Input Auto-Normalize
- Input 200Hz HPF (highpass filter)
- Hum (comb) Filter
- Adaptive Noise Reduction
    - o Predictable Noise Reduction (1CH adaptive filter)
    - o Random Noise Reduction (broadband filter)
- 20-Band Graphic Equalizer
- AGC (automatic gain control)

A 1000-point spectrum analyzer is also provided as an aide in

determining the characteristics of the target audio. Details of each filter are described in the following sections.

3.2 INPUT FILTERS

Figure 2: Input Filters

The two input filters allow pre-conditioning of the audio before it is enhanced by the other filters.

3.2.1 Automatic Normalization Filter

The **Auto-Normalize** filter is used to bring up the level of quiet audio prior to processing. This filter works similarly to the AGC, attempting to bring up the level of the audio to approximately -12 dBFS. The **Auto-Normalize** filter will apply no more than 10 dB of gain, to avoid objectionably increasing low-level background noise. If the input audio level is greater than -12 dBFS, no gain is applied. To enable the **Auto-Normalize** filter, click the LED button so that the indicator "light" turns green.

3.2.2 200 Hz Highpass Filter

The **200 Hz HPF** (Highpass Filter) is used to remove low-frequency noises. For this reason, the input HPF is sometimes called a "rumble filter." The filter removes all sound energy below 200 Hz. Speech energy below 200 Hz rarely contributes to intelligibility, so these low frequencies can be safely removed prior to processing. The filter also removes any DC offset, whether fixed or time-varying, that may be in the audio selection. To enable the **200 Hz HPF** filter click the LED button so that the indicator "light" turns green.

3.3 SPECTRUM ANALYZER

The **Spectrum Analyzer** displays the frequency content of the input and/or output audio. The frequency axis of the **Spectrum Analyzer** plot goes from left to right, with the lowest frequency at the leftmost side and

the highest frequency at the rightmost side. The "energy" (loudness) axis of the **Spectrum Analyzer** plot goes from top to bottom, with stronger (louder) frequencies indicated by higher peaks on the plot.

For example, a signal consisting of a single tone will appear as a single peak, located at the tone frequency. A "white noise" signal, which contains equal amounts of every frequency in the signal bandwidth, will appear as an (approximately) flat line across the entire frequency range.

The spectrum analyzer can display the **Input** audio, the **Output** audio, or both. To enable the **Input** and/or **Output** trace, click the corresponding LED button so that the indicator "light" is on. The **Input** audio is the signal before any filters are applied. The **Output** audio is the signal after all enabled filters are applied. When only one signal (either **Input** or **Output**) is displayed, it is shown in "Bars" mode, where each frequency is indicated with a gradient-colored bar. When both signals are shown, they use "Lines" mode, in which each signal is indicated by a line, using yellow for the **Input** and blue for the **Output**. The "Lines" mode prevents one signal obscuring the other.
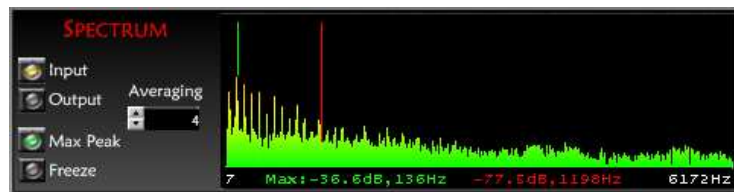

Figure 3: Spectrum Analyzer with Single Channel (Input) Displayed
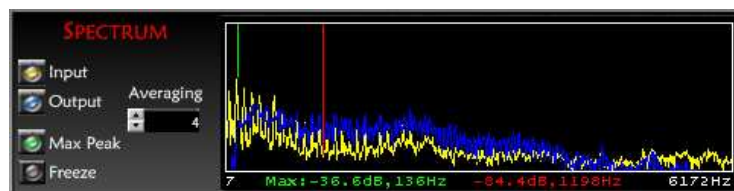
Figure 4: Spectrum Analyzer with Both Input and Output Displayed

The **Spectrum Analyzer** provides a cursor marker to help in identifying specific frequency values. At any time when the **Spectrum Analyzer** is activated, clicking in the graph area displays a red vertical marker at the frequency location clicked. The value (in dB) and frequency (in Hz) of the location clicked is displayed in red text on the bottom right of the graph.

The **Max Peak** feature displays a green vertical line in the graph display at the strongest (loudest) frequency. When this feature is enabled, the maximum peak value (in dB) and frequency (in Hz) is displayed in green text on the bottom left of the graph. To enable this indicator, click

the button beside the **Max Peak** text so that the indicator "light" turns green.

The **Freeze** feature freezes the currently displayed spectrum. This feature serves as a freeze-frame to allow detailed viewing of a specific frequency response at a specific time in the audio. To freeze the currently displayed spectrum, click the **Freeze** LED button so that the indicator "light" turns green. To un-freeze the spectrum display, click the button again so that the indicator light is no longer illuminated.

Use the **Averaging** control to increase or decrease the number of averages applied to the frequency spectrum. A small **Averaging** value allows a more accurate snapshot of the spectrum at a given time, but the trace values will change rapidly, making longer-term spectral characteristics difficult to see. Larger **Averaging** values result in a smoother spectral plot that represents stable frequency characteristics well but does not accurately show rapidly time-varying signal characteristics.

3.4   AUDIO ENHANCEMENT FILTERS

The Audio Enhancement filters provide the main functionality of the QuickEnhance. Each of these filters is designed to remove specific types of noises. These filters can each be used separately or combined to enhance audio that may have multiple noise problems.

3.4.1   Hum Filter

Figure 5: Hum Filter

The **Hum Filter** is used for removing hum noise from the audio selection. (The **Hum Filter** is sometimes referred to as a "comb filter.") A typical source of hum noise is 60 Hz electrical mains hum. This hum is caused by electrical sources such as fluorescent lights and often has strong energy components at multiples of 60 Hz.

The **Hum Filter** fundamental frequency can be adjusted via the arrow buttons or by entering the fundamental frequency in the **Frequency** text box. The fundamental frequency can be adjusted from 40Hz to 75Hz. To

enable the **Hum Filter** click the LED button such that the indicator "light" turns green.

The **Auto-Tracking** feature uses the comb frequency value entered into the **Frequency** text box as a starting point and then attempts to track the fundamental hum frequency present in the input audio. This feature is very useful for detecting hum caused by electrical sources that may not be at the precise 50 or 60 Hz source frequency[1].

To enable this feature, click the LED button beside the **Auto-Tracking** text such that the indicator "light" turns green. When this feature is active, any manual adjustment of the comb frequency turns **Auto-Tracking** off. Disabling the **Auto-Tracking** feature and then re-enabling it resets the **Auto-Tracking** to start at the frequency indicated in the **Frequency** text box. When the feature is first enabled it will take several seconds for the **Auto-Tracking** feature to lock onto the fundamental frequency.

The **Attenuation** slider controls the amount of hum reduction applied by the filter. The maximum hum reduction amount is 60 dB. If the hum filter is objectionably aggressive, try setting the **Attenuation** to a smaller value to reduce the effect of the filter.

3.4.2 Adaptive Noise Reduction Filters

The **Noise Reduction** filters are used to remove both **Predictable** and **Random** type noises. Each of these filters provides an **Attenuation** control for specifying how much filtering is applied.

3.4.2.1 Predictable Noise Reduction
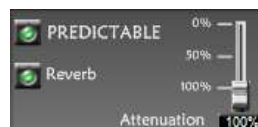
Figure 6: Predictable Noise Reduction Filter

The **Predictable Noise Reduction** filter is used to remove "convolutional noises," or noises which can be predicted based on the immediately preceding audio. Examples of convolutional noises are mains hum, tonal

---

[1] Often, differences in recording speeds can affect the frequencies within the audio track. For example, if when digitizing a cassette the cassette deck plays back the tape at a slightly faster rate than it was recorded at, the resulting digitized audio will have all recorded frequencies shifted slightly higher (faster). Therefore, 60 Hz hum on the original cassette would show up in the digitized recording at a higher frequency such as 60.1 or 61.0 Hz.

components of fan or machine noise, and echoes or reverberation. The **Predictable Noise Reduction** filter is sometimes referred to as a "deconvolver" or a "one-channel adaptive filter." To enable this filter, click the LED button in the upper left corner so that the indicator "light" is green.

When canceling echoes or reverberation, enable the **Reverb** filter configuration by clicking the **Reverb** LED button so that the indicator "light" is green. When the filter is only canceling tones or hum, the **Reverb** mode can be disabled.

Note: The **Predictable Noise Reduction** filter will always cancel both hum/tones and echoes/reverberation to some degree. Enabling the **Reverb** mode allows the filter to better cancel reverberations at the cost of more intense processing and sometimes some additional processing artifacts in the audio. If your audio has only hum/tone type noise, then leave the **Reverb** mode disabled. If your audio has only echo/reverb type noise, or if it has both types of noise, then enable **Reverb** mode.

You can adjust the amount of noise reduction using the **Attenuation** slider. To get maximum reduction, set the slider to 100%. Setting the slider to 0% effectively passes the audio straight through with no change. For most cases, a **Predictable Noise Reduction Attenuation** setting of 100% provides good results.

3.4.2.2 Random Noise Reduction

Figure 7: Random Noise Reduction Filter

The **Random Noise Reduction** filter is used to remove broadband type noises. Examples of broadband noises are wind noise, non-tonal components of fan or machine noise, and background "static" noise. To enable this filter, click the LED button in the upper left corner so that the indicator "light" is green.

The **Aggressive** mode allows the filter to more aggressively attack broadband noises, at the cost of more noticeable artifacts in the resulting audio. Enable **Aggressive** mode by clicking the LED button so that the indicator "light" is green. Leave **Aggressive** mode disabled for steady broadband noises, such as fan noise. Enable **Aggressive** mode if more cancellation is desired for rapidly-varying noises such as wind noise or variable background static.

You can adjust the amount of noise reduction using the **Attenuation** slider. To get maximum noise reduction, set the slider to 100%. Setting the slider to 0% effectively passes the audio straight through with no change. **Random Noise Reduction** introduces a distinctive artifact which, while often less objectionable than the background noise, can be annoying to the listener. The **Attenuation** setting allows you to find the balance point where noise reduction is achieved without introducing too much artifact. Often an **Attenuation** setting in the range 50-75% produces a good result.
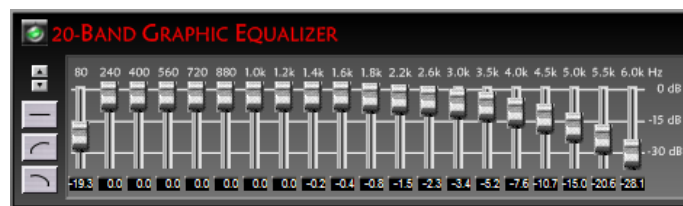
3.4.3   20-Band Graphic Equalizer

Figure 8: 10-Band Graphic Equalizer

The **20-Band Graphic Equalizer** is used to reshape the spectrum of the output audio. The sliders allow attenuation of the audio energy in a series of adjacent frequency bands, with each band centered at the specified frequency. The lower the slider bar is moved, the more attenuation is applied in that frequency band. The attenuation value for each band is displayed beneath the slider, and specific attenuation values may be entered into these text boxes. Valid values are between 0 and 30 dB. To apply the filter, click the LED button so that the indicator "light" turns green.

The buttons to the left of the slider array can be used to quickly adjust the slider array as a whole. The up and down arrow buttons can be used to nudge the overall slider shape up or down. The three "quick select" buttons configure the sliders into a flat, a highpass, or a lowpass shape.

3.5   OUTPUT AGC (AUTOMATIC GAIN CONTROL)

Figure 9: AGC

The **AGC** (Automatic Gain Control) is a dynamic level processor applied to the final output, after all other processing. The **AGC** applies gain as needed to maintain a good audio level (approximately –12 dB). The gain applied by the **AGC** is limited by the **Max Gain** setting; regardless of the audio level, no gain greater than **Max Gain** will be applied. If the signal input to the **AGC** is already –12dB or greater, no gain will be applied.

To apply the filter, click the LED button such that the indicator "light" turns green. You can adjust the maximum amount of gain applied to the audio selection by using the slider or by entering the amount in the **Max Gain** text box.

3.6 REPORT GENERATION

The **Report** button (in the top left corner of the plug-in window) creates and opens a text file named "QEAS_Settings.txt" (for AS version) or "QEVST_Settings.txt" (for VST version) in Windows™ Notepad containing all the current plug-in settings. There is room in this report for the user to enter a project name and his or her initials. The user can chose to save or print the report.

Note that each time the **Report** button is clicked, the "QEAS_Settings.txt" file in the C:\Program Files\QuickEnhance AS II folder will be overwritten. In order to save copies of the report files, be sure to save them to a different file name. The VST version will prompt the user for a new filename each time the **Report** button is clicked.

4 SPECIFICATIONS

4.1 INPUT FILTERS

Auto-Normalize

Normalizes to 25% of full scale (-12 dBFS)

200 Hz HPF

-3dB cutoff frequency at 200Hz, <1dB ripple in the pass-band, >40dB attenuation in the stop-band

4.2 AUDIO ENHANCEMENT FILTERS

Hum Filter

Adjustable fundamental frequency from 40Hz to 75Hz. Harmonic rolloff filter structure reduces notch depth as frequency increases. Auto-tracking feature adjusts to match fundamental hum frequency

Random Noise
Reduction Filter

Adaptive spectral subtraction broadband noise filter.

Predictive Noise
Reduction Filter

LMS adaptive FIR filter of 128 and 1024 taps.

20-Band Graphic EQ

20 band-pass filters centered at 80, 240, 400, 560, 720, 880, 1040, 1200, 1400, 1600, 1800, 2200, 2600, 3000, 3500, 4000, 4500, 5000, 5500, and 6000 Hz. Adjustable from 0 to 30 dB of attenuation

4.3 OUTPUT AGC

AGC

No level adjustment necessary. Automatically adapts to loud or quiet signals to ensure approximately –12 dBFS average audio level, subject to an adjustable maximum gain. Adjustable maximum gain ranges from 0 to 60 dB

4.4 OTHER FEATURES

Bandwidth

7500 Hz

Specifications subject to change without notice.

5 AUDIO NOISE REDUCTION PRIMER

This noise reduction primer is presented in an informal, easy-to-read manner to help you better understand the characteristics of sound, noise, and speech. The information given here is intended for those unfamiliar with audio clarification. For more thorough training in audio processing, DAC teaches DAC School courses that include lecture presentations and hands-on experiments.

5.1 SOUND

Was that a sound?

Sound is simply pressure waves in the air that are caused by some disturbance (such as a tree falling, a car crashing, or a person talking). In other words, something happens that causes the air to respond in a certain way that affects our ears and causes us to sense sound. At 70° F (21° C), these sound waves travel through the air at about 1130 feet per second (344 meters per second) or about 770 miles per hour. The rule of thumb we typically use is that sound moves at approximately one thousand feet per second or one foot per millisecond.

Sound has two characteristics that are important to us when we are clarifying audio: energy and frequency. Energy is the strength of the signal, but we sometimes refer to it as level or (somewhat incorrectly) volume. We often measure the audio level in decibels (dB). High dB is loud, while low dB is quiet. Decibels are measured on a special logarithmic scale which enables us to use a practical range to indicate both very loud sounds and very soft sounds. Below are some common sounds and their dB level.

Level (dBA SPL)	Example
135	threshold of pain
125	jack hammer
115	car horn
95	subway train
75	street traffic
65	conversation
55	business office
45	living room
35	library reading room
25	bedroom at night
15	broadcast studio
5	threshold of hearing

Table 1: Typical Sound Pressure Levels

Frequency is a characteristic that describes the repetition of the disturbance over time. We measure frequency in units of hertz (Hz), which indicate the number of cycles per second (cps). The frequency of sound affects our ability to hear it. Remember that our ears respond to this disturbance in air. A good ear typically responds to frequencies between 20 and 20,000 Hz. As we age, or if we have damaged ears, the higher frequencies usually become less audible. Law enforcement officers who years ago took target practice without good-quality hearing protection frequently demonstrate loss of high frequency hearing. (This is not entirely bad for audio clarification, as we will explain in the section on speech.)

In summary, sound has two defining characteristics: energy (or level) and frequency. In later sections we will discuss what we can do with these characteristics to concentrate on speech clarification.

5.2 NOISE

Man, do we have a noisy world!

Now that we understand sound, let's talk about a special kind of sound — noise. For our purposes, anything but the voice signal in which we are interested is classified as noise. There is a lot of noise around: in the house, at the office, in traffic, at the park...everywhere. In some places they even add noise so that you don't notice other noises. We are constantly exposed to noise. Most of the time, it just makes our good tape recordings a little bit harder to understand, but sometimes the noises are so loud that we wonder whether there are actually voices on the tape!

We classify noises into three categories: additive, convolutional, and distortion. The sketch below shows a typical recording scenario and some of these noises. We will discuss each of the three categories separately.
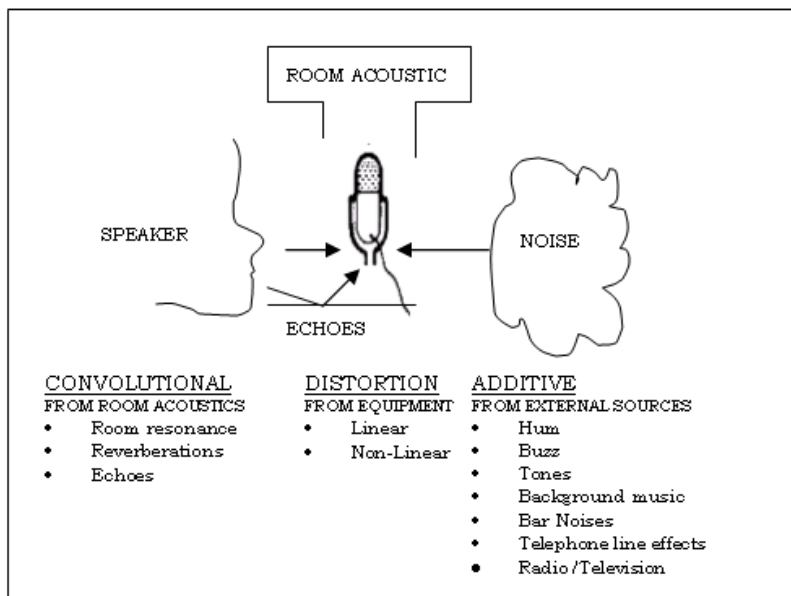

Figure 9: Noise Model

5.2.1 Additive

The most common of the noises we encounter are additive. These noises are anything that generates an audible sound. The list is endless: engines, motors, fans, florescent light buzz, radios, TV, juke box, music bands, glass breaking, paper rustling, wind, etc., etc., etc.. For audio clarification purposes, these additive noises can be classified into two general categories: random and time-correlated. Certainly you know

what random noise is like. Wind noise comes to mind. Also sounds such as a nylon jacket rubbing against a "concealed" microphone. *Sound familiar?* Many noises we encounter are classified as random. Time-correlated noises, though, have a repeatable, predictable "pattern." Examples are tones, power line hum, and fluorescent light buzz. These noises can be effectively reduced using deconvolver filtering, and often this processing alone is enough to make a recording intelligible, even in the presence of random noises that cannot be reduced by the process.

5.2.2 Convolutional

The second type of noise is convolutional. These noises are the result of room acoustics and only exist if there is a sound source. In a room with no noise source, there is no convolutional noise. If a person speaks or some additive noise source is present, depending on the room acoustics, there may be an echo or reverberation. This echo is a convolutional noise. Sometimes it is so strong that it interferes with hearing the desired audio signal. It is also a repeatable, predictable signal (described above) and thus a time-correlated noise. In "hard (reverberant) rooms," for example (sheet rock walls and ceiling, tile floor, no soft furniture, nothing to absorb sound waves), a sensitive microphone will pick up echoes that our ears might ignore, making the recording much worse than expected. Jail cells and interview rooms are good examples of hard rooms, and are the source of many bad recordings encountered by law enforcement. But, remember echoes are time-correlated, and deconvolver filtering can reduce time-correlated noises. Again this processing is often sufficient to allow an otherwise unintelligible recording to be understood.

5.2.3 Distortion

The third type of noise is distortion. The equipment we use to make the recording introduces this noise. Inexpensive microphones, bad cabling, poor quality recorders, cheap tapes, weak batteries — all contribute to distortion. Distortion, though, cannot be reduced without physically altering the recorded sounds. This could raise questions of "doctoring" if we tried to use the altered recording in court. We are lucky, though, because distortion typically has only a minimal effect on voice intelligibility; the main impact is on voice quality. Since we cannot do much with it, we will not address distortion any further, except to say that we should pay attention to all of the components of our system. Remember the old cliché about the weakest link — it applies here, too.

Now putting noise into our "sound" perspective, it, too, can be defined in terms of frequency or frequency range (bandwidth) and energy level. And we will discover that some are easily reduced while others are not affected by current audio clarification techniques.

5.3   SPEECH

What a nice conversation to listen to!

For the purpose of this primer, we are going to concentrate on trying to reduce noise to hear the speech. In some applications, there may be other compelling reasons to analyze the entire audio signal, but for our purposes we will just try to make the voices more intelligible. We are interested not in analyzing the voice but only in determining what is said.

The human voice communicates an abundance of information of which words are but a small part. The speaker's sex and age, emotional state, level of education, dialectical influences, and physical attributes are all communicated in the acoustic waveform. In order to preserve all of this information, the voice should be recorded and processed with as much fidelity as possible.

Now remember, we said that many people can hear audio signals up to 20,000 Hz. That's quite a range, but for our purposes much of that frequency range consists primarily of noise. Pay attention here: the most important speech information is located between 200 and 5000 Hz! In most cases, audio signals outside this range will be considered "noise" according to our definition. Sometimes a voice will have frequency characteristics outside this range (low or high) so it may be necessary to adjust our processing frequency band. Remember these numbers, as they will continually come into play when we begin our process of reducing background noise. So ***as a rule of thumb***, we can establish the **voice frequency range as 200 – 5000 Hz.**

Vowels typically have high energy concentrated below 3000 Hz. Consonants, on the other hand, have lower energy and typically are distributed both below and above 3000 Hz. Both are very important for word discrimination. Speech is also classified as a semi-random audio signal because its components are relatively unpredictable. The figure below shows the average spectrum of many people's voices.
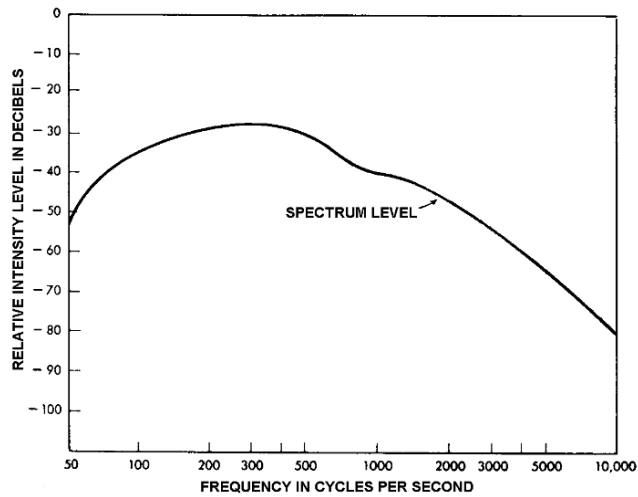
Figure 10: Typical Averaged Voice Spectrum

The figure above shows that there is voice information out to 10,000 Hz and beyond. However, when we expand the upper frequency limit, we find that the signal usually contains substantially more noise than voice information. We use the term signal-to-noise ratio, S/N or SNR, to indicate the relative amount of energy in the desired signal compared with the interfering noise. The signal-to-noise ratio is defined as:

$$S/N = \frac{signal\ energy}{noise\ energy}$$

If we expand the bandwidth, it will only marginally increase the signal (speech) energy but will substantially increase the noise level. The S/N ratio decreases as a result, and it is harder to understand the voices. We want the highest signal-to-noise ratio possible when we are trying to understand what is being said. We have found that voice quality and intelligibility are acceptable for a voice frequency band of 200 to 5000 Hz. Expanding beyond these limits has marginal value.

5.4 MONITORING AND RECORDING

I'm glad we got that on tape!

This section is a brief discussion on recorders, primarily on how they impact audio enhancement. We will cover three types of recorders that are commonly used in law enforcement: microcassette, cassette and digital recorders.

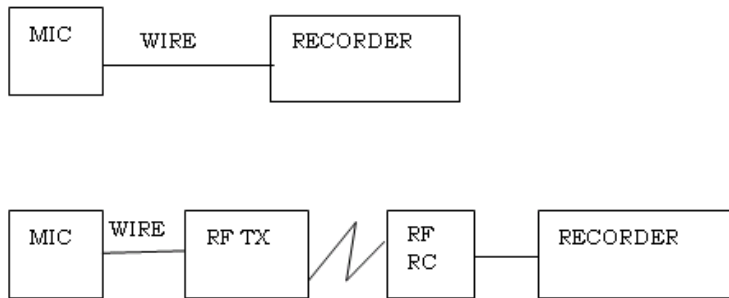Typical recording setups are shown in the block diagrams below:

Figure 11: Typical Recording Setups

The quality of each component is critical to the quality of the tape recording. Surely this is no surprise. It is like your stereo system at home. If you have the best amplifier and best CD player but cheap speakers, what kind of sound can you expect? Similarly, if you want a good recording, you have to pick up the sound, transmit it with good quality equipment, and then have a good quality recorder.

5.4.1 Microcassette tapes

The microcassette is commonly used in law enforcement primarily because of its size. There are two major draw backs to obtaining a good quality recording with this device: first, these units are typically bandlimited to around 3000 Hz; and second, the quality of the tape is generally poor. The microcassette design has very little dynamic range above 3000 Hz, which results in reduced frequency response. Remember, there is important voice information up to 5000 Hz. If the frequency bandwidth of the recorder is too low, the recorded voice may not sound like the person speaking. If you use microcassettes, it is recommended that you get a high-quality microcassette playback machine which has pitch control to account for varying tape speeds and azimuth adjustment which allows you to match the playback head to the recorded tracks. It is also recommended that that you not use the built-in microphone as these are generally poor quality, and they are typically not located at an optimal position for recording purposes.

5.4.2 Cassette tapes

Cassette recorders are still the mainstay of law enforcement audio recording. There are many good-quality units available. Again, the built-in microphone should be used only as a last resort. These recorders have bandwidths to 10,000 Hz, which is more than adequate for recording voice. The highest recordable frequency is determined by two principal factors, tape speed and head gap. The faster the speed and the narrower the gap, the higher the recordable bandwidth. Many recorders have been modified to reduce the speed by one-half in order to allow twice as much time to be recorded onto tape. You do not get something for nothing: changing design speeds results in reducing the highest potentially recordable frequency to lower than the desirable 5000 Hz.

Another problem with analog magnetic tape players is dynamic frequency variations referred to as wow and flutter. These effects are due to the tape's passing over the record/playback head at a varying speed. Motor speed regulation, varying tape tension, and irregularities in pinch and backup roller shape all contribute. Wow is a low frequency variation (a few Hz), while flutter is a high frequency variation (up to hundreds of Hz). Wow and flutter are generated during both the recording and the playback processes. The overall effect of the fluctuations is to produce an undesired "modulation" effect on the recorded audio. Substantial wow and flutter are audible as a vibrating, or nervous, overtone to the voice, accompanied by a loss of audio crispness. Even modest levels of wow and flutter impair clarification and noise cancellation because the signal processor is forced to "chase" the noise.

Although it is not always possible, tape playback should be done on the same machine that did the recording. This can help minimize some of the deficiencies of the specific recording equipment .

5.4.3 Digital recorders

Digital recorders (e.g., DATs, MiniDiscs, recordable CDs, and SSABRs) are becoming more popular as their price decreases. These recording units offer exceptional quality for audio reproducibility. The incoming analog signal is converted into a sequence of numbers in an analog-to-digital (A/D) converter. The audio is sampled at rates up to 48,000 samples per second (and beyond) to define the signal precisely, with no introduction of wow or flutter distortion. These samples are stored as binary numbers (1 or 0). The data can be recorded onto magnetic tapes, floppy disks, hard disks, compact discs, DVDs, or flash memory chips. Using a laser, the data can be optically recorded. Even if the recorded data has very poor noise characteristics, it is still possible to distinguish between 1 and 0 to achieve a good quality playback.

A word of caution here when using digital recorders. Some digital recorders employ <u>lossy audio compression</u>, but lossy compression hampers audio clarification. Consider this example: if a chip can store 1,000,000 bits of data, and the recorder has 10,000 bits of audio per minute real time, the recorder can store 100 minutes of data. Lossy compression throws away some audio data that is deemed by the algorithm to be unimportant, allowing the chip to record more than 100 minutes of audio. Lossless compression looks for redundancies but does not actually throw any data away. Depending on the compression algorithm, our ability to subsequently reduce or cancel noise from the recorded signal may be affected. It is therefore recommended that any audio compression feature used be a lossless compression or that no compression be used when making digital recordings that might require subsequent processing.

5.4.4 General recording techniques

Remember that the recording, or even live monitoring, is only as good as the weakest component. In addition, the equipment should be cleaned, examined for damage, and tested prior to each use. Always use fresh batteries and a new, good quality tape. Whenever possible, we should make stereo recordings because processing techniques can use stereo to further reduce background interference.

Microphone placement is critical to making a good recording. You want as clear an air path to the mic as possible. The spacing between stereo microphones should be about the size of your fist with the thumb extended (approximating the distance between the ears). The audio level is inversely proportional to the square of the distance to the microphone. In other words, if the distance from the sound source to the microphone is doubled, the audio level will be cut to one-fourth the original level. Suppose a person being recorded is five feet from the microphone, and his voice has a sound pressure of 20 pascals at the mic. If he moves away to ten feet from the microphone, the sound pressure drops to 5 pascals. (Because decibels use a logarithmic scale, the sound pressure level would drop 12 dB from 60 dB to 48 dB.) A man's voice at a long distance from the microphone becomes a low frequency, muddled tone blending into the background.

For body wires, always fasten the microphone to the outermost piece of clothing, fastening in such a way to prevent the clothing from rubbing against the mic. For room mics, place them close to where the people are actually talking and away from strong noise sources. Good recordings are hard enough to obtain without our making matters worse by failing to take these reasonable steps before the recording is made. Also, by taking these steps, we ensure that we will get the best possible results when we go back and process the recorded audio to reduce noises.

5.5   AUDIO ENHANCEMENT

Now that sounds better!

Now we are getting to the technology that DAC does best. Despite all of our efforts to use good equipment and properly plan for the recording of the conversations, background noise (over which we have little control) often makes voices difficult to hear and understand.

There are a variety of audio filters that we can use to reduce noise. Our goal is to understand what is being said. But if we also plan to use this tape in any legal proceedings, we must be careful that the voice sounds like the person recorded. This means that we must be careful what we do in the voice frequency range.

5.5.1   Set bandwidth

The first step in our enhancement process is to establish the frequency bandwidth. All DAC equipment either has adjustable bandwidth or is preset to cover the voice frequency range. Remember that microcassette recorders cannot record audio above about 3000 Hz, so processing at a higher bandwidth is unnecessary. DAT recorders can record up to 20,000 Hz, but important voice information is generally located at 5000 Hz and below. There is no reason to process signals outside this range as they are mostly noise. So first determine the highest voice frequency recorded, and then set the equipment bandwidth accordingly. QuickEnhance has a preset bandwidth of 8 kHz.

Since we are discussing bandwidth, let us discuss a special type of law enforcement recording where the signal is automatically bandlimited. A recording of a telephone conversation obtained by wiretap has the signal bandlimited to about 3200 Hz by the equipment at the telephone company. Higher frequencies, which might make the voices more distinguishable, and easier to understand, simply <u>are not present</u>. Perhaps you have noticed that it is difficult to recognize the voices of certain people when they call you on the phone; the loss of the frequency information beyond 3200 Hz is the primary reason. Thus, when we process a recording of a telephone conversation, we are forced to set the upper frequency limit close to 3200 Hz so as not to process excess noise.

5.5.2   Apply audio filters

After setting the bandwidth as close as possible to the voice frequency range (200 – 5000 Hz) or to the equipment-limited bandwidth (e.g., 3200 Hz for telephone), we are ready to apply additional audio filters to reduce more background noises. The most common audio filters are the high-

pass, low-pass, notch, and comb filters. The most important filters we will use will be the adaptive predictive deconvolver (Predictable Noise Reduction, or the one-channel adaptive filter), and the broadband (Random) noise reduction filter. We will also cover the use of a digital graphic equalizer.

5.5.2.1 Highpass filter

The high-pass filter, sometimes called a rumble filter, reduces noise below a specified cutoff frequency. If we set the filter's cutoff frequency at 200 Hz, all signal energy below 200 Hz will be reduced according to the specified stop-band attenuation, thereby making these frequencies much less audible.

5.5.2.2 Lowpass filter

The low-pass filter, sometimes referred to as a hiss filter, reduces all signals above a specified cutoff frequency. If we set the filter's cutoff frequency at 4000 Hz, all signal energy above 4000 Hz will be reduced according to the specified stop-band attenuation, making these frequencies much less audible.

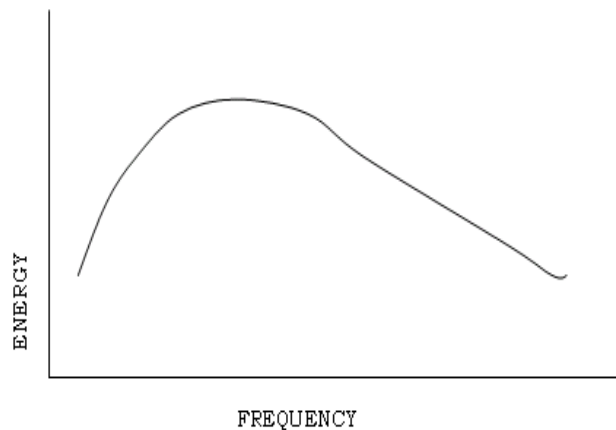To illustrate the highpass and lowpass filters, look at the figures below:
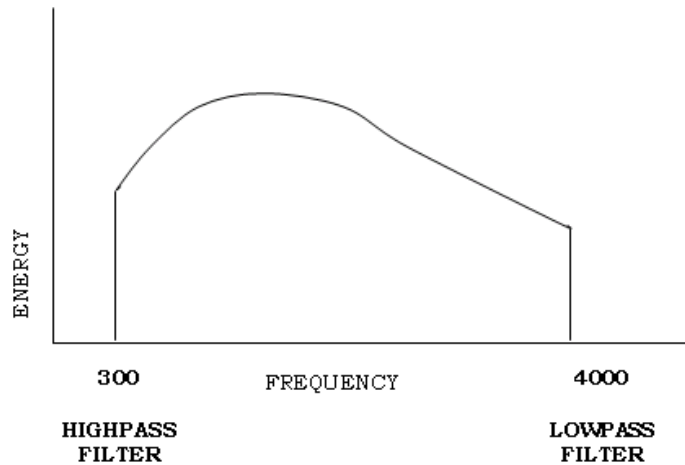
Figure 12: Input Signal

Figure 13: Processed Signal

5.5.2.3 Notch filter

If we are able to identify a narrow frequency band in which a noise exists, we can apply a notch filter that will affect only that signal. For example, if we set the notch frequency at 1000 Hz and the notch width at 200 Hz, the narrow band of energy between 900 and 1100 Hz can be reduced according to the specified notch depth.
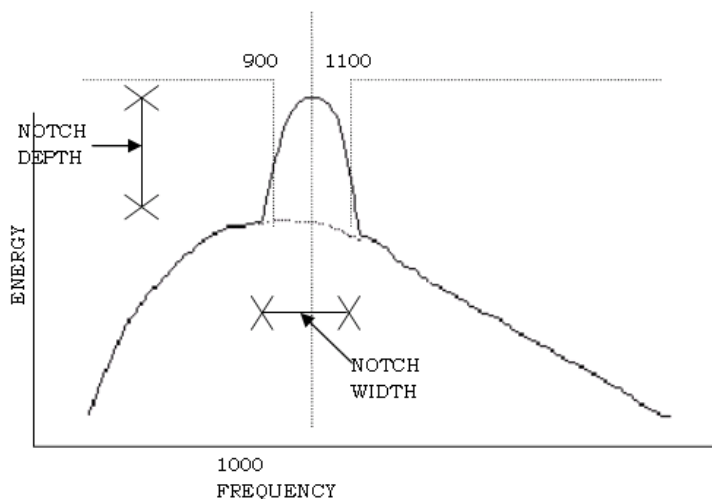

Figure 14: Notch Filter

We must be careful, though, when using a notch filter. If the notch width is set too wide, or the depth is set too deep, the quality of the voice will be affected. It may not "sound" like the person speaking.

5.5.2.4 Comb filter

The comb filter is designed specifically to reduce power line hum (60 Hz in U.S. and 50 Hz overseas), and its harmonics. (Harmonics are simply multiples of the fundamental frequency; such as 120 Hz, 180 Hz, 240 Hz, etc.) Sometimes when our equipment is not grounded properly or is located too close to an AC power line, a "hum" occurs in the audio signal. A comb filter acts similar to a notch filter by reducing energy at both the appropriate frequency and its harmonics. The depth of a comb filter is constant for all harmonics. Some people prefer to use multiple notch filters where the depth can be adjusted at each harmonic.
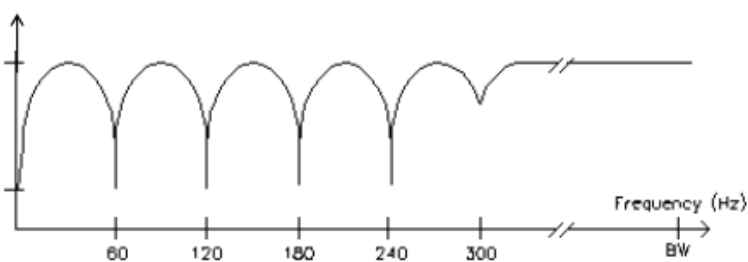

Figure 15: Comb Filter

5.5.2.5 Other bandlimiting filters

There are three other lesser-used audio filters that are available in some DAC products for special application as described below:

- Band-pass filter – a combination of a high-pass and low-pass filter where the lower and upper cutoff frequencies can be specified.
- Band-stop filter – the opposite of the band-pass filter that passes audio below the lower cutoff and above the upper cutoff frequencies.
- Slot filter – the opposite of the notch filter where only the audio within the slot width is passed.

5.5.2.6 Adaptive filter

The **one-channel (1CH) adaptive filter** is sometimes referred to as an *automatic noise reduction filter* or *deconvolver*. The other filters we have discussed are called fixed filters; that is, when we set the parameters, they do not change unless we change them. The adaptive filters, on the other hand, constantly analyze the incoming audio signal and automatically make adjustments when they detect changes in the noise. No operator action is required.

It would be nice if this filter reduced <u>all</u> background noises, but unfortunately this is not the case. The filter only reduces repeatable or predictable signals. We call these <u>time-correlated</u> signals, and several of the forms of noise mentioned earlier are time-correlated. Hums, tones, echoes, and reverberations are examples of time-correlated signals. The assumption is that any time a signal is time-correlated (I've heard it before and expect to hear it again), it is not voice. Remember that speech is a relatively random audio signal. Reducing all time-correlated signals has little impact on the voice and usually does a very good job of making the voice more intelligible. Unfortunately, though, many noises in nature are random (water/shower running, wind noise, etc.); these signals are unaffected by the 1CH adaptive filter. Some sounds, such as a fan or motor, have both periodic and random components, and a 1CH adaptive filter can remove the periodic sound.

The adaptive filter has two major advantages: first, it automatically adjusts itself when the noise changes; and second, it can operate in the voice frequency bandwidth to reduce noise and not affect the voice. This makes the adaptive filter ideal for use in unattended voice surveillance activities.

5.5.2.7 Equalizer

An equalizer is also a common type of audio processor. Many home stereos even include an equalizer. DAC includes a multi-band digital graphic equalizer in several of its products. Simplistically, this equalizer consists of multiple bandpass filters, with each band covering a portion of the total bandwidth. As each slide control is moved, the energy level in that frequency band is reduced or increased. This can be used to reduce noise levels but is more often used to reshape the voice spectrum to improve voice quality. Remember the typical voice spectrum? After processing, the resultant spectrum may be more flat. By using the equalizer as shown below, the voice can be returned to a more typical spectrum, thus improving voice quality.

5.5.3 Audio clarification procedures

The primary purpose of reducing background noise is to be able to hear what is being said – voice intelligibility. If we plan to use the enhanced tape in court or other legal proceedings, the voice on the tape must also "sound" like the person. This could limit what we can do to process the audio. Remember, the adaptive filter can reduce time-correlated noise within the voice frequency bandwidth and not affect the voice. But what if there are random noises still present which mask the conversation? If these noise signals extend across the entire bandwidth (broadband noise), there is little we can do. However, if they are limited in their frequency range(s), we might be able to do something.

Suppose that there is some random noise between 3000–3500 Hz that is so loud we can not hear the conversation. We could use either a notch filter or a band-stop filter to reduce the loud noise in this frequency band. The result may be that the noise level is reduced and we can hear what is being said. However, the voice energy would be modified within this frequency band so it probably would not "sound" the same as the person normally does. Thus, in this case we would have to find the proper balance when adjusting the filters so that we achieve both maximum intelligibility and maximum voice quality.

In conclusion, this primer has provided an introduction to audio enhancement and noise filtering for beginners. There are other filters and many techniques and processes that were not covered here. To learn more details on audio enhancement, you are encouraged to attend DAC School.