AUDIO Postproduction for Digital Video

Digital Video DV EXPERT SERIES

CMP*Books*

AUDIO CD INCLUDED

by Jay Rose

AUDIO POSTPRODUCTION for DIGITAL VIDEO

Jay Rose

Digital Video DV EXPERT SERIES

CMPBooks
San Francisco, CA • New York, NY • Lawrence, KS

Published by CMP Books
an imprint of CMP Media LLC
Main office: 600 Harrison Street, San Francisco, CA 94107 USA
Tel: 415-947-6615; fax: 415-947-6015
Editorial office: 1601 West 23rd Street, Suite 200, Lawrence, KS 66046 USA
www.cmpbooks.com
email: books@cmp.com

Technical editor:	David Moulton
Acquisitions editor:	Dorothy Cox
Copyeditor:	Lydia Linker
Managing editor and layout design:	Michelle O'Neal
Cover layout design:	Damien Castaneda

Distributed to the book trade in the U.S. by:
Publishers Group West
1700 Fourth Street
Berkeley, CA 94710
1-800-788-3123

Distributed in Canada by:
Jaguar Book Group
100 Armstrong Avenue
Georgetown, Ontario M6K 3E7 Canada
905-877-4483

For individual orders and for information on special discounts for quantity orders, please contact:
CMP Books Distribution Center, 6600 Silacci Way, Gilroy, CA 95020
Tel: 1-800-500-6875 or 408-848-3854; fax: 408-848-5784
email: cmp@rushorder.com; Web: www.cmpbooks.com

Printed in the United States of America

02 03 04 05 06 5 4 3 2 1

ISBN: 1-57820-116-0

CMP **Books**

Table of Contents

Chapter 16

Noise Reduction

Remember this:

- You can often lower the level of perfectly steady, pitched noises (such as lamp dimmer buzz or camera whine) dramatically, and in some cases eliminate them entirely.
- You usually can't get rid of random or unpitched noises (such as traffic or preamp noise) without compromising dialog. The best you can do is make them less annoying.
- Equalizers, expanders, and delays can also be effective noise reducers. Instructions and examples are in previous chapters.

It's Not Really Noise Reduction

Today's sophisticated software and expensive DSP-driven boxes do an amazing job of distinguishing random noise from speech and applying sneaky techniques to make the noise less objectionable.

But make no mistake; if random noise occurs at the same time as dialog, in the same frequency band, you can't get rid of it. At least, not with today's technology.[1] Sorry. No amount of marketing hype, wishful thinking, or impressive prepackaged demos at trade shows will change that fact.

This isn't to say you can't improve most noisy tracks, even on the desktop, by using the right processes intelligently. You can certainly lower noise level in places where it doesn't compete with dialog. But you can only hope the words are loud enough to distract viewers from the noise where it does. If noise isn't too serious and the dialog is otherwise well-recorded, this strategy can be remarkably successful.

1. I can see a future device that would use speech recognition and a highly evolved version of the vocal modeling we used in Chapter 15, to synthesize tiny chunks of dialog when noise obscures it. But we're not there yet.

That's what this chapter is about. But you've got to understand the ground rules, and one of the first is that random noise during dialog can never be truly eliminated.

The Reality of Noise Removal

A graphic analogy can help you understand why random noise never really goes away.

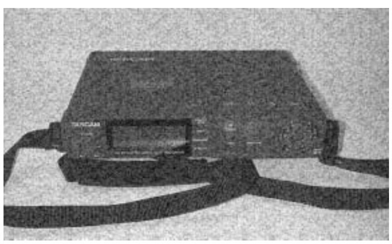
> **Gotcha**
>
> ***But isn't all noise random?*** I'm using random to mean the sound is of indeterminate frequency, such as electronic hisses, traffic, generator noise, and mumbling production assistants.
>
> Non-random noise, in a technical sense, has absolutely steady frequenices. This includes pure tones and even constant, harmonically rich sounds like dimmer buzz or ground-loop hum. If it's not too loud compared to dialog, this kind of noise may be removable.

Let's assume the DAT recorder in Figure 16.1 is the visual equivalent of well-recorded dialog.[2] Constant random noise, such as the hiss from recording at too low a level, could be equivalent to the dirty gray pattern laid over it in Figure 16.2. This noise is too random to remove with an equalizer or filter any other way because such a filter would also affect the dialog.

But a graphic editing program, properly tuned, *can* tell the difference between the recorder and its background. Likewise, noise-reduction software, properly tuned, *can* tell the difference between dialog and noise. It lowers the volume when dialog stops, the equivalent of our lightening the non-recorder areas in Figure 16.3. As you can see, the noise over the recorder hasn't changed at all—and we've also lost a lot of shadow variation in the background.

The audio noise gate—for decades, this has been just about the only tool we had for this purpose—works exactly like the graphic one in the figures. It's off when there's no dialog, and it turns on when it hears sound above a preset threshold. Usually there's a floor control which lets

16.1 This clean photo could be an analogy for well-recorded dialog.



16.2 And this dirty one could be the analogy for a noisy track.

16.3 Noise reduction can make the overall picture look better, but the dirt is still there.

2. Aptly enough, it's a Tascam DA-P1, a model frequently used for double-system sound at DV shoots.

Gotcha

Is noise reduction just for dialog? Current pop music is too broadband for noise reduction software to do an effective job. Other forms of complex music that have some dynamics and spectral movement may benefit from noise reduction if it was badly recorded. But the algorithms can also destroy some musical details, so use them carefully.

Field-recorded sound effects often benefit from noise reduction. This is particularly true of staccato and moderately soft effects such as footsteps, or loud effects with long reverb tails such as gunshots.

We'll refer to dialog noise reduction in this chapter, simply because that's the most common use for the process. The principles are the same no matter what you're trying to clean up.

a small amount of the background noise through at all times—which would account for the reduced but not erased shadow information in Figure 16.3.

Today's noise reduction software is a lot more sophisticated. Systems use hundreds of upward expanders tuned to narrow bands across the spectrum. Noise is allowed through only at those frequencies where there happens to be voice energy. At the same time, noise in other bands is attenuated. Thanks to a quirk in how our hearing mechanism works, this can trick the ear into thinking the noise is completely gone. It's still there; we just don't hear it.

Masking

Our ears are not particularly precise sensors. While each ear has close to 30,000 nerves on the basilar membrane and while these nerves are tuned to respond to different pitches, there isn't a single nerve for each possible frequency in hertz.[3] Neural information gets combined in various ways before it reaches the brain.

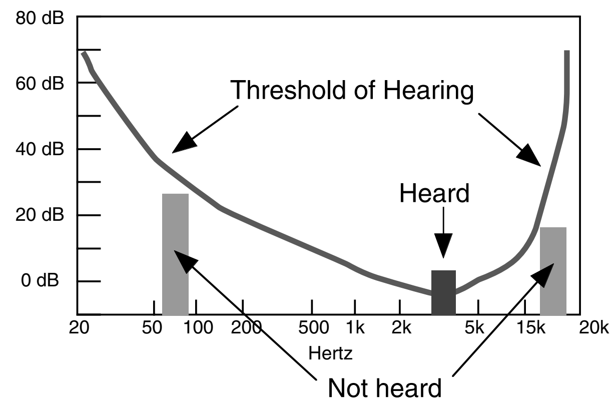
When we hear a tone at a particular frequency, a group of nerves centered around that pitch fire. How many nerves will go off depends on the volume of the tone as well as other factors. A loud sound triggers more nerves. These combined nerves are interpreted as a single pitch at a certain volume. But because that loud sound involved a bunch of nerves around its frequency, softer sounds at a nearby frequencies might not be able get through—the nerves or neural pathway that should respond to them are already doing other things.

This phenomenon has been known for years and has been measured across very large populations. It affects the threshold of hearing. The heavy gray line in Figure 16.4 represents that threshold. The decibels are calibrated relative to the frequency where most people's ears are the most sensitive, around 3.5 kHz. You could consider 0 dB on this chart to be true 0 dB SPL—the nominal threshold of hearing—or any other convenient level, depending on the individual.

3. How could there be? The ear's design predates the scientific concept of hertz by at least a couple of years.

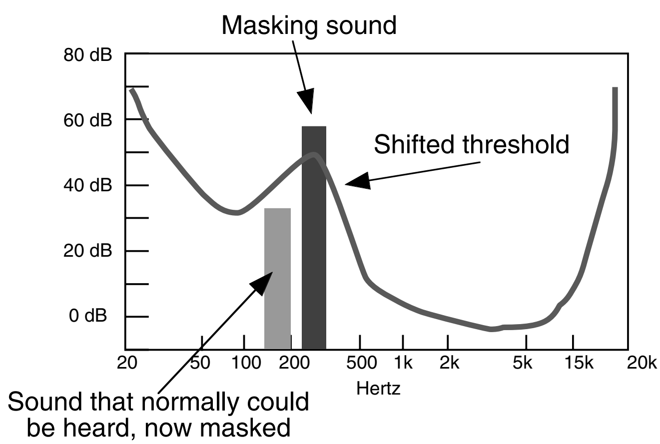
The important thing isn't how the vertical axis is calibrated; it's what happens between the center of the band and the extremes. At 3.5 kHz, the short, dark gray bar is louder than the threshold, and it gets heard. But at 50 Hz or 15 kHz, most people won't detect a sound until it gets 40 dB louder. Even though the light gray bars are taller and much louder than the dark one, these light grey ones represent sounds that would get lost.

Unfortunately, that heavy gray line isn't fixed. When something sufficiently loud comes along (dark gray bar in Figure 16.5), it drags the threshold with it. A 250 Hz sound, 25 dB above the threshold, ties up so much neural activity that a simultaneous 200 Hz sound 10 dB softer (light gray bar) isn't heard. The actual amount of masking varies with the frequency, volume, and overall timbre of both sounds, but it's always there to some degree.



16.4 A typical threshold of hearing curve. Sounds below the gray line are lost for most people.

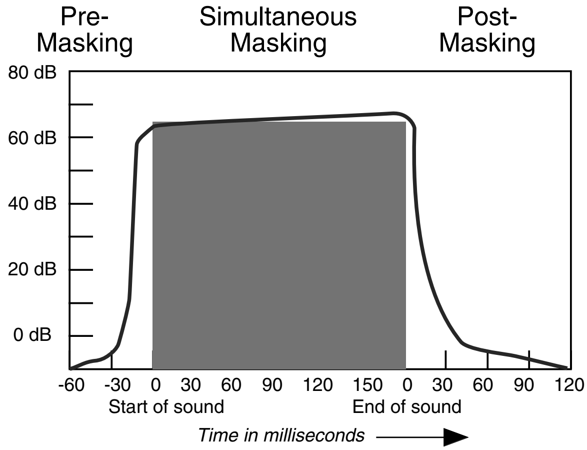
A similar effect occurs over time, both because it takes a moment for the brain to recognize sounds and because nerves have to recover after being fired. While this effect also varies for different sounds, Figure 16.6 shows a typical *temporal* masking. In this example soft sounds, between a dozen milliseconds before to up to 50 ms after, are masked.



16.5 Frequency-based masking at work. The louder sound moves the threshold above the softer sound.

So if we can arrange to have noise only at times and frequencies where it'll be masked by dialog, the noise effectively goes away. A high-precision spectrum analyzer would still see the noise, but we'd never hear it.

Masking, by the way, is also the secret behind perceptual encoders like mp3 and AAC. When used properly, these algorithms can shrink an audio file's size by 90 percent with no apparent audible effect,[4] or shrink it even more with minor losses in quality.



16.6 Temporal masking means you can miss sounds that occur a short time before or after a louder sound at a nearby frequency.

Masking is also a reason why it's important not to have elements of a track compete in spectrum or in time. In general, the closer two sounds are in frequency and the farther in volume, the more the softer one will be masked. Masking usually starts when sounds at similar frequencies are within about 10 dB. Loud, low-frequency sounds often mask a wider range than high-frequency ones of the same volume.

A Few Other Noise Reduction Facts

There are plenty of myths and misconceptions about the process of noise reduction. If you understand what's really going on, you'll usually get better results.

Noise Reduction Without a Noise Reducer

Noise reduction software often attempts to take care of things automatically. This can do more damage, in many cases, than using other techniques manually. If you've been following the examples on this book's CD, you've already heard some fairly effective noise reduction that relies on other kinds of processors.

- Get rid of whistles using the equalization techniques in Chapter 12 and demonstrated on Track 51.
- Tracks 54 and 55 show how downward expansion can improve modestly noisy interview tracks (Chapter 13).

4. Some people swear they can always hear *any* encoding. But auditory studies show this probably isn't true. mp3 has gotten a bad rep, primarily because of some awful files on the Web, poorly designed encoders, and people who don't know how to use the technology properly. You'll learn how do it right—and hear how transparent good encoding can be—in Chapter 19.

- Track 62 virtually eliminates dimmer buzz using a comb filter (Chapter 14).

And of course, the ultimate noise reduction for extremely bad dialog recordings is ADR (Chapter 8). These tools are often the first defense, and should be considered before whipping out general-purpose noise reduction software.

Editing can also be used for noise reduction: replace the noisy part with something else. That's one of the principal uses for room tone, though noise-reduction editing can be as subtle as replacing a few waves with clones of adjacent ones. It can even involve changing individual samples by drawing over clicks or other transient sounds with the pencil tool in an audio editing program. A few high-priced DSP-based noise reducers can do these things automatically, though they need precise tuning to sound good.

Nulling noise?

Folks who know a little acoustic theory sometimes ask, "Why can't I create or capture a sample of the noise, invert the polarity, and use that to cancel the noise out?". The idea behind this is basically correct: the comb filter, for example, works by delaying symmetrical noise exactly one-half cycle. This lines up the negative side of the wave against the positive, *nulling* out the noise. But that's a special case, relying on a characteristic of the noise itself and using a precise delay instead of a capture, or copy and paste.

For the sample-and-invert technique to work, the noise would have to be of absolutely consistent waveform and volume, and the sample would have to be pasted back exactly in sync with the continuing noise. We don't have any way to do that with today's technology. If there is the tiniest error in timing or if the noise has changed since the sample, this operation would increase the noise instead of removing it.

There are some noise reduction processes that rely on a sample of the noise, but they're not using an inverted sample. Instead, they take a spectral fingerprint of the noise and use that to control other techniques. Nulling isn't involved.

Dolby noise reduction

The Dolby process revolutionized analog recording, and it is still used in many studios. But it only reduces noise part of the time—just enough to make noisy transmission channels seem quieter.

Dolby A—the original format—worked by breaking the audio into four bands. Each band was then compressed, raising the volume of its soft sounds and decreasing the dynamic range. The four signals were combined and recorded on tape at a high volume. Analog tape adds hiss to a recording. But on playback, the Dolby signal was again split into bands. Each was downward expanded, restoring the dynamic range and—during average or soft passages—lowering the amount of hiss. The combination of compression during record and expansion during playback, is known as *companding*.

During loud passages, no compression or expansion took place. The hiss remained at its usual level, but it was soft enough to be masked by the loud signal. Dolby A used four bands to help the masking; in a single-band system, a loud bass note would momentarily reduce the expansion, letting us hear unmasked hiss in the mids and highs. A newer Dolby format—*Dolby SR*—continuously adjusts the bands based on audio content, helping masking even more. It's used today by music producers who consider analog tape an important step in creating a unique sound.

Consumer Dolby (*B* and *C*), popular in analog cassette decks, also split the signal. But it passed the low frequencies unchanged and only companded the highs. That's because hiss is more of a problem at high frequencies on smaller, slower tape formats. Both versions of consumer Dolby required careful calibration so the compressor and expander would precisely match. If calibration slipped (something common in consumer decks), the spectral balance would be changed. A competing scheme, *dbx*, used a single band with a different companding model that didn't need calibration.

This kind of noise reduction was considered *double-ended* because it used equipment during both recording and playback, specifically to keep tape noise from interfering. Neither system did anything for electronic noise from prior operations or that may have been picked up by the mic.[5] Both Dolby and dbx have since moved on to other things. Current Dolby digital systems don't rely on companding, and dbx is now in the business of making studio compressors and equalizers, as well as single-ended noise reducers.

Double-ended noise reduction isn't necessary in digital production. All of the techniques in this book are single-ended.

Noise Reduction Examples

I can't provide a cookbook for this chapter because every noisy recording has its own unique problems. Instead, I'm providing a few examples of real field recordings, showing the basic steps to clean them and playing the results. Use these as a benchmark for your own efforts.

Hear for yourself

Track 69 plays before-and-after sounds for the noise reduction example.

5. Dolby A's four-band decoder could be applied to an unencoded noisy signal and, with careful adjustment of levels, function as a single-ended noise reducer. This use was popular for a while on film mixing stages. But we've got much better tools now.

Gotcha

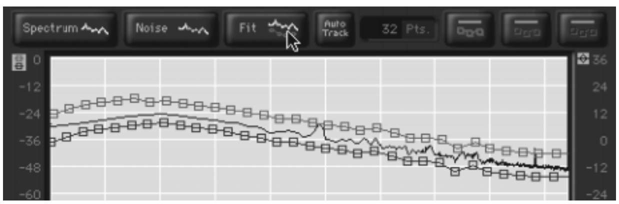
There's a lot more in the arsenal. If you've jumped to this section to solve a particular problem and haven't read the rest of this chapter, you may be missing some major noise-reduction techniques.

Check Chapters 12–14 to learn about noise reduction using equalizers, expanders, and delays. In many cases, these can remove noise with fewer artifacts than specific noise-reduction software.

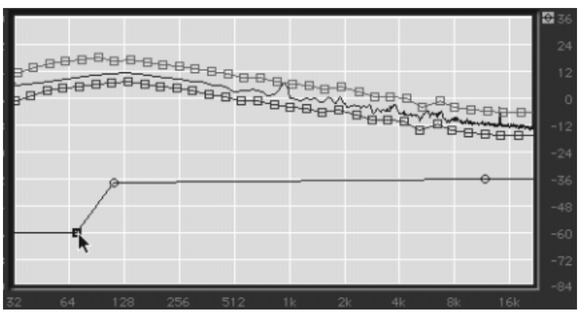
Moderate Interior Noise

Part 1 of Track 69 is the original track of an interior interview. There's some room rumble and HVAC hum, which we'll remove in Ionizer.

16.12 The first step is to analyze a short sample of the noise.



16.13 Step 2: Fitting 32 expanders around the noise.



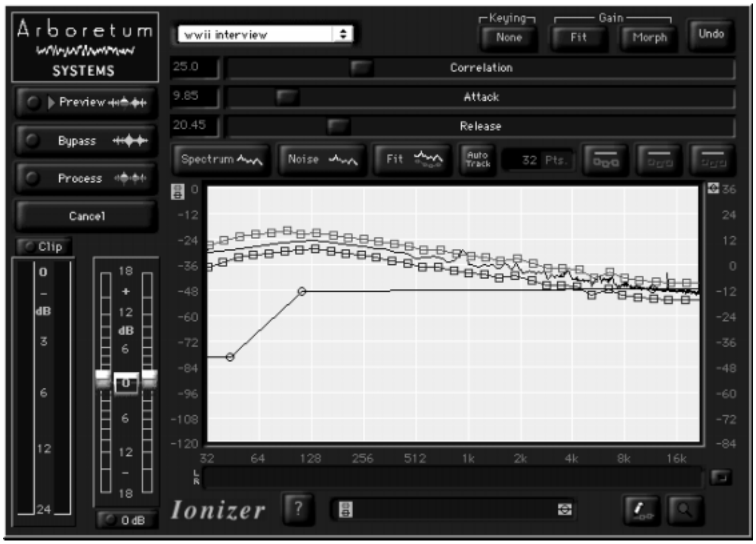
16.14 Step 3: Drawing a floor for the expanders.

The first step is to select a short slice of the noise during a pause. Look for a place that doesn't include breathing or other noises that don't occur during the rest of the track. Make sure you haven't included any of the subject's voice. In this case, I found about 20 frames around 12 seconds from the start.

Open Ionizer and press the Spectrum button. The software quickly analyzes the noise you selected, creating a graph of level versus frequency as in Figure 16.12.

Then cancel Ionizer—it remembers its settings after you close—and select the whole clip. Open Ionizer again and press the Fit button to create 32 downward expanders around the noise's spectrum. This looks like Figure 16.13. Each square is a potential control point. The top line shows the expanders' thresholds; they're just above the noise so that anything softer than the threshold will be reduced in level. The bottom line is where the expanders have a full cutoff. The area between the lines is the expanders' knees. I used only 32 relatively wide bands, even though the software is capable of 512 narrow ones, to avoid filter distortion and other artifacts.

Then draw a floor for the expander. In Ionizer, the numbers to the right of the graph (Figure 16.14) show attenuation. In that figure, there's a floor of -36 dB for most of the noise, but it goes considerably deeper below 100 Hz to reduce rumble. Part 2 of Track 69 plays the result—the noise is almost completely gone. In fact, it was too clean and sounded unnatural.



16.15 The final settings to clean up the interview.

Raise the floor somewhat. Figure 16.15 shows the final settings, including correlation and time constants. Part 3 plays the result.

Constant, Loud, Mechanical Noise

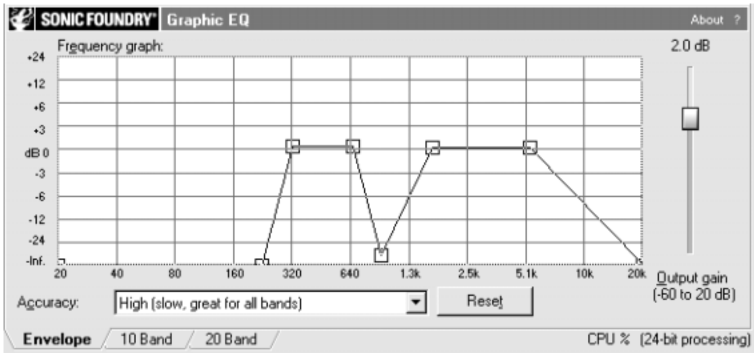
Hear for yourself

Track 70 is the before-and-after for motorboat interview example.

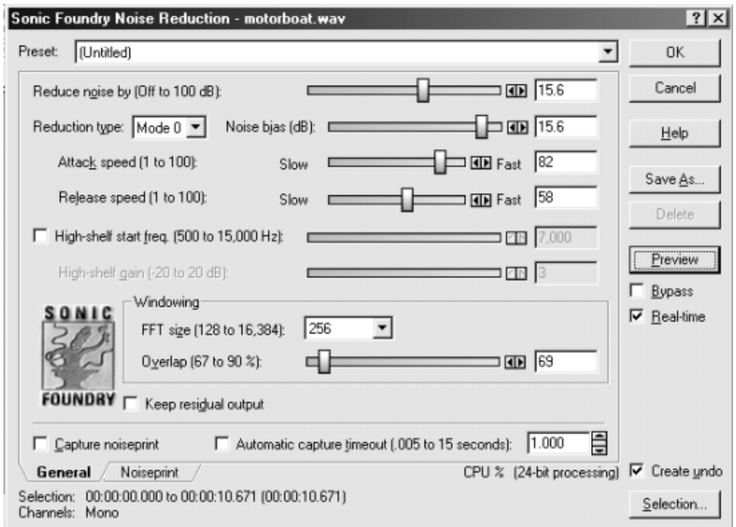
Part 1 of Track 70 is an interview recorded on a motorboat. The motor is almost as loud as the voices and is at many of the same frequencies. The best you can hope for in a situation like this is to improve the intelligibility. (Besides, people looking at the shot can tell it's on a moving motorboat, so the noise won't seem out of place.) I decided to process it in SoundForge, using Sonic Foundry's separate Noise Reduction plug-in.

The first step is to help things along by knocking down noise at frequencies where the voices aren't very active. In theory, this shouldn't do much at all—the motor isn't competing with the voices there. But in practice, lowering the overall noise helps viewers concentrate on the dialog. So apply an equalizer which draws the graph in Figure 16.16. The low-frequency cut hurts the voice fundamentals, but the motor is too darned loud there; I decided the formants would be sufficient. The midrange dip is because spoken voices are typically weak there. But of course, you can turn Preview on and do the tweaking by ear (as you should always tune *any* equalizer).

Then apply noise reduction. Sonic Foundry's is very similar to Ionizer, and the process is the same; sample some noise by itself, fit the expanders to it, select the entire clip, and process. The sequence is slightly different and you don't leave the plug-in to select the whole clip; click the plug-in's Help button for details. Figure 16.17 shows the settings; part 2 of Track 70 plays the result. Because extreme processing was necessary, the voice is left with an echoey artifact. But it's much better than the unprocessed file.



16.16 The background was just too noisy to process directly, so the first step was to equalize it.



16.17 These are the Noise Reduction settings for the equalized file.

Moderate Exterior Noise with Whine

If the noise has any constant-pitch elements, pre-equalization will always make noise reduction work better. Part 1 of Track 71 is an exterior interview with fairly loud traffic and a machine whine. The whine competes with, and is almost as loud as, the subject's voice; standard noise reduction techniques won't be able to do much.

Hear for yourself

Track 71 is the whiny exterior before processing, using just a noise reducer, and using a parametric equalizer before the noise reducer.

Part 2 plays the interview after being processed with Ionizer, but with no separate equalization. It's an improvement, but there's still noise. Worse, the expander makes the whine pulsate in step with the voice.

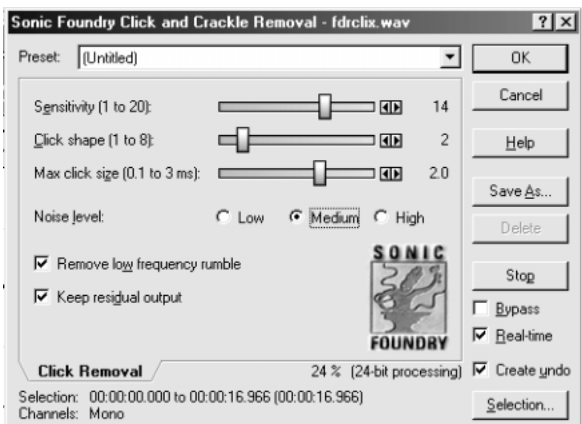
In part 3, the unprocessed interview is run through a parametric first, tuned to eliminate the whine (using the sweeping technique in Chapter 12). Then it's analyzed and noise-reduced by Ionizer. The whine isn't part of the noise print any more, so the spectrum doesn't have big jumps. This means that expanders are less likely to interact with a flangey sound, and I can lower the floor. Listen to all three tracks. Compared to part 3, part 2 seems hardly noise-reduced at all.

Click Reduction

Part 1 of Track 72 is a badly damaged 78 RPM phonograph recording, originally recorded in 1933—probably a lot worse than anything you're likely to encounter. Figure 16.18 shows the settings to clean it up in Sonic Foundry's Click and Crackle remover. This is a powerful plug-in, with lots of options. Read its online Help file; the settings in that figure wouldn't be best for a more modern recording.

Hear for yourself

Track 72 is a segment from Franklin Delano Roosevelt's first Inaugural Speech. This 70-year-old recording has plenty of clicks, which are easily removed in part 2 of the track.



16.18 Declicking a particularly bad 78 RPM in Sonic Foundry's Click and Crackle remover.

Part 2 plays the result. Clicks, particularly during pauses, are seriously reduced. They're also reduced during speech, but because so much processing was necessary, they're often replaced by distortion. However, without those high-frequency clicks, you can apply some creative equalization to improve intelligibility. I'll leave that as an exercise for you.