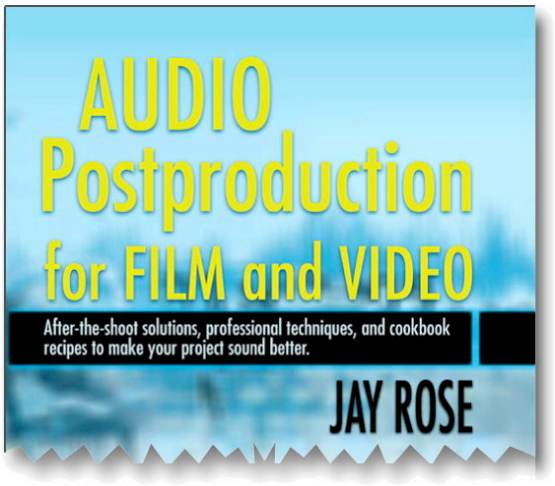

Audio Postproduction
for Film and Video, 2nd Edition

Jay Rose, CAS

Focal Press • October 2008
ISBN 9780240809717
http://www.dplay.com/book/app2e

This is unformatted. The actual book pages are professionally designed, and the illustrations are bigger.

Excerpts from

Chapter 8: Editing Dialog

From page 159...

This chapter is in three sections

The first section is what dialog editing is really about: the ability to patch individual words—or even tiny syllables—together to make a seamless whole. It's how you avoid noises or mispronunciations in an otherwise perfect take, how you combine two less-than-perfect takes to make a better final result, and how you tighten up an interview to keep the meaning but make a better documentary or commercial.

For this book, I'm considering dialog to be any time a human talks: scripted lines, on-the-street interviews, voice-overs and narrations, animation voices, historic clips. Film theorists might break these into smaller categories, and they're usually treated differently when it's time to mix, but when you're editing it's mostly all the same.

The key word is "seamless". Properly cut dialog doesn't sound edited at all. It sounds like the person talking said exactly the right thing. When dialog editing isn't seamless—when you can hear a click, or awkward jump in pitch, or strange pacing—it momentarily kills the message. It reminds you that what you're watching has been manipulated.

If you've ever cut a film or video, you're aware of the need for smooth dialog. But if you've ever watched a good dialog editor, you were probably amazed that all these seamless edits happened quickly, without trial and error or multiple undos. The first section of this chapter teaches you how to work that way: read it, and you'll know exactly where to place the cursor, and which tiny sounds work together when others won't. While it's based on speech science, it's grounded in reality and actually pretty easy. I've taught hundreds of picture editors to work this way.

The second section has to do with workflow, and what happens to a track between picture edit and the mix. These evolved differently for film and video. Each has advantages and drawbacks. But the historic differences don't exist anymore: You can use both techniques on the same project, and accomplish a lot more.

There's a third section, at the end, on restoring lost sync for dialog.

From page 161...

We'll start with a quick review of the most basic technique, the one taught in NLE manuals. Track 19 of the book's CD contains audio from a typical narration and a talking-head sequence. Let's start with the narration, recorded in a quiet studio. The woman is saying three sentences; we'll take out the middle one. If you want, load the first

part of Track 19 into your NLE and follow along—though this example is so trivial, you'll understand it easily from the pictures.

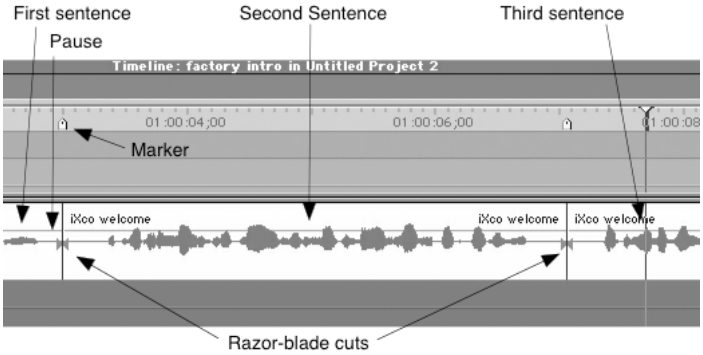

Figure 8.1
A basic, almost brainless edit of a narrator recorded in a quiet studio.

1. Put the clip on a timeline, set so you can see its waveform.
2. Visually locate the two pauses between the three sentences.
3. Put the cursor in the middle of the first pause and plant a marker. Repeat for the second pause.
4. Using the razor tool, cut at the two markers. It should now look like Figure 8.1.[1]
5. Delete the middle section and slide the other two together. Or do a ripple edit, dragging the second razor cut over the first.

This edit works fine if the background is quiet. But the method starts to fall apart when there are constantly changing background noises, such as traffic or machines. That's because there's a good chance the noise will change abruptly at the edit point.

> *Hear for yourself*
> *Track 19 contains source audio for this example and the next.*

Coping with location noise

Unlike the previous exercise, which can be understood from the figure, this one is worth actually doing. Load the second part of Track 19 (a location interview in a moderately noisy medical lab) into your editor. Cut out the middle sentence using the steps above. The razor cuts should look similar to Figure 8.2.

Figure 8.2
Razor cuts in a location interview. The background noise will cause problems.

After you do the deletion or ripple edit, look closely at the cut (Figure 8.3). That tiny change in the background noise, exactly at the edit point, can sound significant. Listen to it in your version or play the first part of Track 20.

[1] This picture is from Final Cut Pro. But it should look about the same in most NLEs and many multi-track audio programs.

Figure 8.3

You can hear room tone jump at this edit point.

You could apply a long cross-fade, but that takes time to render and still might not sound natural. In fact, as a general rule:

> **Rule**
>
> **Cross-fades longer than a half frame have very little place when joining dialog clips[2]. They can be useful in track splitting, but that's a different technique.**

From page 170...

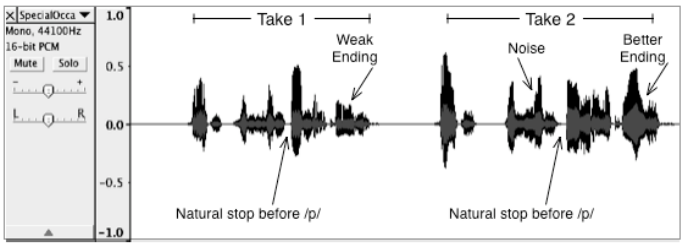

Figure 8.9

Two takes: The first is cleaner... but flatter. Read how to combine the two seamlessly.

The line was delivered at normal speed and there's no pause between "a" and "special," so you can't cut in silence between those words. But the /p/ in "special" is a stop consonant; it creates a pause *inside* the word. You can scrub through the two takes, easily hear and mark that pause in each take, and edit from one to the other: "This is a sp || ecial occasion". The whole process shouldn't take more than ten seconds and should sound perfect.

This principle also works when the words aren't strictly identical. If you had the right words from elsewhere in the clip, you could join them to the first phrase to say, "This is a spectacular show," or "This is a specious argument."

Phoneme-based Editing Rules[1]

Cutting the way we just did is fast, precise, and usually undetectable. First, hear the phrase you want to edit, slowly, in your head. Identify any phonemes that might be useful for the cut you want to make, using the rules below. Then scrub through the clip to find those points.

All the *stop consonants* are created by storing air pressure in your mouth and then releasing it in a burst. There's a moment of silence in the middle of each stop consonant, right before the pressure is released. It can be as short as a third of a frame.

> **Rule**
>
> **Stop consonants will always give you a moment of silence.**
>
> **If a stop consonant is followed by a pause, it usually has two distinct sounds: one when the pressure is cut off, and another**

[1] It really *does*, man. But what I meant was, "Here are some rules for using phonemes when you edit."

when it's released. But the second part isn't important and can be deleted to shorten the pause, or edited to some other word.

If two stop consonants are next to each other (as in, "fat cat"), they're usually *elided*; the closure comes from the first consonant, and the release from the second. But when people are self-conscious, th[1]ey often pronounce each stop separately, making four distinct sounds. Editing out the middle two stops will make a nervous speaker ("the faT Tcat") sound more relaxed ("the fah..Tcat").

With the exception of /h/, *friction consonants* are created by forcing air through a narrow opening: between the lips for /f/ and /v/, between the tip of the tongue and back of the teeth for /th/ and /TH/, and so on. This always makes a high-pitched sound that's easy to spot while scrubbing.

From page 177...

Track splitting and filling

A rare thing happened in Hollywood a number of years ago: something originally done to save money actually improved the art. You know how scenes are shot: A dialog line might extend from a master shot to a two-shot to a close-up, get a response from a character in a different close-up, and then go to another close-up or back to the master. Eventually, all these little bits of picture have to be cut together. What we see as a continuous scene might have had three or four different boom mic angles, performances, and background noises. Unless something is done in audio post, they'll all sound different on the screen.

TV sound doesn't worry about this. The medium grew up with continuous performances, shot with multiple cameras and switched in real-time. It made audio post a lot easier... if there was any at all (remember, TV started as a live medium). As videos moved out of real-time studios and began to be assembled in editing systems, the aesthetic and philosophy stayed. Video editors would assemble single dialog tracks on their masters, and audio post would do things to sweeten them, usually at the same time as adding music and sound effects.

Part of this also had to do with playback. Until recently, video usually stayed on small screens with small speakers. Movies, of course, are heard through big speakers. Any sudden shifts in voice timbre or noise level jump right out, destroying the mood. Since films are shot and edited in pieces, there are dozens of shifts in each scene. It takes good monitors and proper processors to fix them. But film dub stages are big and expensive, so studio bosses didn't want to take any longer than necessary in those rooms.

The answer was brilliant: Take the film editor's single track, and split it into pieces depending on what needs to be done.

Splitting: Why and How

Figure 8.11 shows how dialog fixes typically happen in TV production. A single audio track arrives from the picture editor, drawn here as a piece of moving tape with waveforms. Most of this scene's waveform is of a consistent quality (area A on the tape). Once the console has been set up, it can be pretty much ignored; the engineer can easily tweak overall dialog levels while mixing the ambience and music.

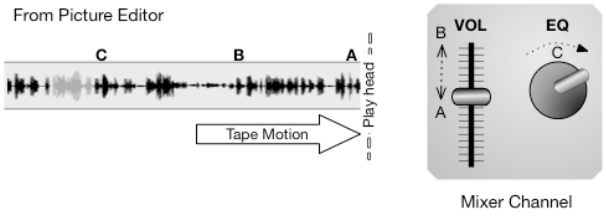

Figure 8.11: A typical TV dialog workflow. The jumps at B and C can be fixed easily by tweaking level or equalization.

On this particular tape, the area starting at B is too soft. And C has the wrong timbre (drawn as a non-matching gray). But the engineer has plenty of time to raise the level just for B, then grab the equalization knob in time to fix C.

Now look at Figure 8.12, a typical scene of the same length from a film production. Because it was built from multiple takes, there are many more shifts in quality. And since it'll be heard on giant speakers, smaller changes in timbre are more obvious. That's why they're shown in two different grays.

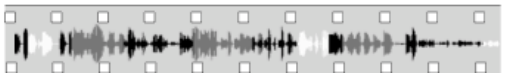

Figure 9/12: A typical film scene has a lot more variation.

The film's re-recording engineer might be able to catch all those changes on-the-fly, but it's harder. It will take a lot of rehearsals or automation programming to get all the cues and adjustments just right. Since a whole movie's dialog is usually made up this way, this could slow mixing down to a crawl.

Instead, the film goes to a *dialog editor* first. That editor splits the picture department's one track into four separate ones, sorted by sound quality and separated by silence (Figure 8.13). Each needs its own kind of processing. But the engineer can preset four console channels with the proper settings for each, rewind the whole scene, and let it rip.

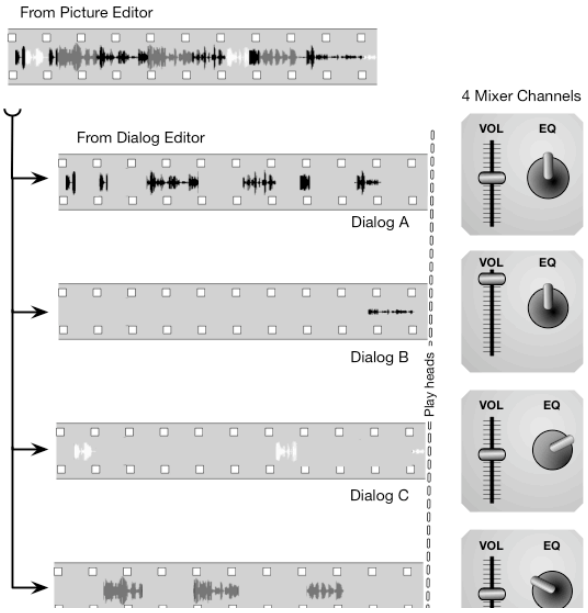

Figure 8.13

Splitting the film into multiple dialog tracks will make the mix a lot easier[1]. All four tracks are run simultaneously, so a piece of the original dialog is playing at all times.

1. The actual process involves separating pieces of the original dialog with silent, non-magnetic leader. I left those splices out for clarity.

From page 184...

Restoring Lipsync by Eye

It also helps to make your editing setup as sync-friendly as possible. Use a large picture monitor and play at full frame rate. Make sure playback is smooth, and be sure to allow enough time for preroll and synchronization if your system requires it. The best software environments give you *bump* or *nudge* controls—single key commands that let you slip selected regions against picture in controlled increments of a frame or less, while both are playing.

There's a general procedure to make sync matching a little easier.

1. Play the audio clip and locate the first plosive /b/ or /p/ sound. Find the silence right before that sound—it can last a frame or two, and unless there's a lot of background noise you'll be able to see it on the waveform—and place a marker on the first frame where the sound returns.
2. Note approximately how far that marker is from the start of the speech.
3. Find the first frame of video where the subject's mouth opens.
4. Move forward the same distance you noted in step 2, less a few frames.
5. Shuttle forward slowly until you see the lips close. Then jog forward and find the first frame where the lips start to open again and put a marker there.

That step requires a good eye because the movement can be subtle, but it's always there. Figure 8.15 shows a typical /p/: note how the lips are just starting to part at 1:07:20:11, after having been closed in the previous frame.

Figure 8.15: Find where pursed lips just start to open after being closed, and you've located a /b/ or /p/.

6. Line the two markers up and things should be in perfect sync. If they look slightly off, move the sound a frame forward and check again. If that makes it look worse, go back the other way. If moving helps but doesn't fix things, add another frame in the same direction. If sync didn't look right at all when you first checked, you found the wrong lip movement in step 5. Try again.

If the speech doesn't have a /b/ or /p/ sound near the front, listen for an /m/ followed by a vowel. Follow the same steps—lip movements for an /m/ looks almost the same as those for a /p/—but mark the audio clip on the start of the vowel sound after the /m/

You may be able to skip steps 2 and 4, and just match the first sound with the first picture where the mouth is open. Sometimes this works, but often it doesn't. I've found it faster to always go for the stop consonant.

This material is excerpted from
Audio Postproduction for Film & Video, 2nd edition.
Full details at http://www.dplay.com/book/app2e