

Article: SONAR 4 Vocal Production Workshop
by Rick Paul - 8th July 2005 -



Most of our articles here on CakewalkNet tend to focus on tools that can help you get hot instrumental sounds and grooves going in your SONAR projects. It goes without saying that your talent and imagination are far more important than the specific instruments, loops, and processors you use. However, given whatever talent level you might have, odds are that using higher quality tools can help improve your results.

If your SONAR projects are recordings of songs, though, all the softsynths and loops in the world can only go so far toward helping you achieve top quality results. When it comes to songs, the quality of the vocals is often the most critical component of the recording. Even the hottest instrumental tracks will have a hard time compensating for a mediocre vocal performance because we humans tend to identify much more readily with the human voice than we do with instrumental sounds.

It probably goes without saying that the most important ingredient for achieving super vocal recordings is a great singer. The caliber of microphone and mic preamp you are using also comes into play. For those of us with our own major record label and/or unlimited budgets, there may be no worries in any of these areas. For the rest of us, though, we may have to make do with whatever it is we have at any given time. If whatever we have results in vocal recordings that do not measure up to the sound we hear in our heads or on our favorite CDs, are we simply out of luck?

It is tough to make up for a less-than-stellar singer. However, if the major labels can take less-than-stellar, but video-friendly, singers and polish their performances such that we'd never know they weren't great singers, at least until hearing them live, why can't we take whatever level of vocal talent we have available and apply some of the same tricks and techniques to improve our own vocal recordings? For that matter, even great singers mess up from time to time, and, if your window of time with the singer has long since passed, it may be necessary to work with whatever you have in order to turn out the best possible recording.

This article will focus on techniques SONAR 4 users can use to take whatever level of raw vocal recordings they may have and make them the best they can be. Whether that means achieving polished radio-ready results, or just ratcheting things up a notch, will depend on what you have available and how much time and energy you want to expend trying to improve things. Some of the techniques discussed will involve third party plug-ins, but all techniques discussed can be performed within SONAR 4. (SONAR 4.0.2 was used for validation of the techniques for this article.)

What Goes In

We will mostly be talking about techniques you can use after the vocal performance has been tracked. While what goes on before the recorded vocal actually gets into SONAR is probably more important than most, if not all, of what comes afterward -- after all, the singer's performance is paramount -- most of that will be beyond our control. Besides, if you're interested in more information on microphone and preamp selection, vocal techniques, DAW architecture, and so on, there are numerous places to find that. For example, it seems like most of the major recording magazines run primers on mic selection every year or so. However, for those who may be new to vocal

recording, it is worth spending a few minutes outlining the main components of our input signal chain. Those who have more experience may also be curious as to what assumptions are being made in this area that may affect what is discussed in the rest of the article.

It is tempting to think of the input signal chain as beginning with the microphone since that is the first piece of gear we, as recording engineers, can control. However, the first element in the vocal recording chain is actually the most important element -- i.e. the singer. The room in which the singer is recorded also enters the picture prior to the microphone since any reverberations of the singer's voice in the room may color the recorded signal. Everything else we'll be talking about anywhere down the line, from the microphone on, will be geared at making the singer sound good. Things are a whole lot easier, however, if the singer is already sounding good in the room.

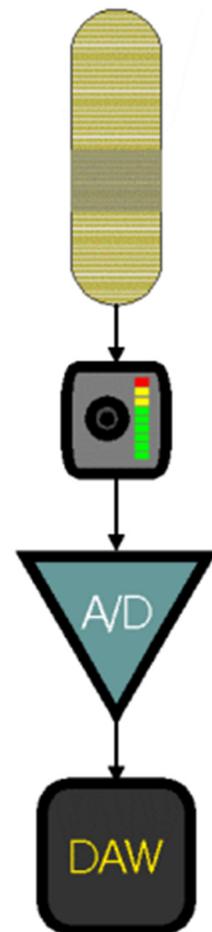
This isn't just a "garbage in, garbage out" consideration, though. Rather, it is to point out something that may not be as obvious to individuals who haven't recorded singers before or who haven't recorded more than one singer. The singer (and, to a lesser degree, the room in which the singer is being recorded) is the most unique element in the signal chain. Many people will have the same microphone, preamp, DAW, plug-ins, etc., but there is only one of the singer you are recording. In fact, it is more extreme than that because that singer may sound very different from one day to the next, or even from one hour to the next, as environmental conditions, wear on the voice, personal comfort issues, mood, and more combine to affect the singer's performance.

This uniqueness means there isn't a "one size fits all" approach to recording vocals along the lines of, "use a Brand X (expensive tube) mic, through a Brand Y (expensive tube) preamplifier, through a Brand Z (megabucks classic) compressor, ..." While the basic building blocks of the signal chain may be similar from singer to singer and song to song, you will need to listen to the singer in his or her natural state, to identify what is special about his or her voice, and what that implies in terms of gear and processing to flatter the singer and suit the song.

That expensive Brand X mic might be something that works great for a number of big name singers, but might be horrible with your singer's voice. Short of having years of experience in listening to voices and microphones and being able to match them up intuitively, the only way to figure out what is right is to listen and experiment. Just remember that it's about the singer, not the gear. That said, let's move on to the gear. (I know you fellow gearaholics were just chomping at the bit.)

First up is the microphone. The mic is responsible for capturing the singer's voice in the room and converting it to electrical signals. It goes without saying that not all mics are created equal. It may not be as obvious that a "better" mic (typically meaning a "more expensive" one) is not necessarily better for all singers, or even any given singer on all songs. We're not going to get into mic selection, but it is worth mentioning that you would ideally select whatever mic best suits the combination of singer and song so what goes in is as good as it can be. If you have a limited mic closet, though -- I've only got two mics myself, neither of them of the expensive vintage type -- you may not have so many options, and we'll get to a possible after-tracking solution later in this article.

The signal coming out of the microphone is generally too low for the uses we intend making of it later on, so the next element in the signal chain is the mic preamp, which is responsible for boosting the mic's signal to usable levels. Here again, the ideal scenario would be matching the preamp to the singer, song, and mic, and you may or may not have multiple options available to you. We aren't going to worry about that here (though you should for your own recordings). There are two things we will mention regarding preamps, though. One is that preamps come in many forms, with some key forms being standalone preamps, preamps built into computer audio interfaces, and preamps built into mixers. Of particular note is that, if you are using a preamp built into your audio interface, it may be physically housed in the same box as the next component in line, the analog-to-digital (A/D) converter. The other is that, while you are using the preamp to boost the mic's signal, one important consideration is that you do not boost it to the point where your A/D converter overloads and clips the incoming signal. This could cause nasty digital distortion



in the recorded signal that cannot be repaired later on.

Due to this second consideration, some may choose to place a compressor and/or limiter between the mic preamp and the analog-to-digital converter. I have not put a compressor or limiter in the signal chain diagram above because it is optional, and I prefer to avoid doing that. The reason is twofold: First, I prefer to retain as much flexibility as possible for later on, such as at mix time, and using a compressor while tracking commits you to the sound and settings of that compressor. Secondly, the one hardware compressor/limiter I have available is not of as high quality as the plug-in compressors I have for use inside SONAR. Thus, I set preamp levels conservatively, to avoid the possibility of A/D converter overload, and rely on using 24-bit audio to give me plenty of resolution, even at low volume levels.

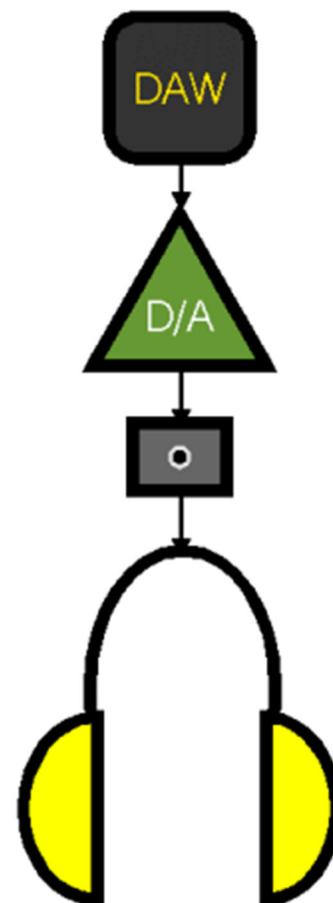
Speaking of A/D converters, this is where the analog electrical signal coming in from the mic preamp or any compressor/limiter after the preamp gets converted into digital audio data. For many home studios, the A/D converter will be built into the audio interface. However it could be in a standalone unit, part of a preamp with a digital output, or even part of a device designed for some other function which just happens to include A/D converters (e.g. I once used an ADAT for A/D converters, with a digital optical link from that to the sound card I had at the time, which did not have its own A/D converters). Settings associated with this stage will govern the bit depth (e.g. 16-bit or 24-bit) of the audio signal coming into SONAR, as well as the sample rate (e.g. 44.1kHz, 48kHz, 88.2kHz, 96kHz, 192kHz) to be used. Since SONAR currently only supports one bit-depth and sample rate per project, these will need to be consistent with the rest of your recording project, and are more likely to be governed by your preferences and project delivery needs than by anything related to vocal quality. However, if you are limited to using 16 bits as your bit depth, you may need to consider using a compressor and/or limiter prior to the A/D converter to deal with the limited resolution of the digital audio while avoiding overloading the A/D converter.

In our simple block diagram above, the next step in the input chain is to pass the digital signal from the A/D converters to your DAW (i.e. SONAR) for recording and other processing. While there may be other steps between the A/D converters and SONAR -- e.g. if your sound card supports onboard processing, you may be able to place some of that processing between the A/D converters and the handoff of the data to SONAR -- the most common scenario is to preserve this simple input chain. While there may be advantages to using some onboard processing, namely a compressor and/or limiter, prior to the A/D converter, I see little reason for placing one after the A/D converter and prior to the recorded signal. Doing so would have all the disadvantages of tracking with effects without the benefit of being able to optimize the levels of what gets digitized. Any other processing needed can be added after the unaltered signal is recorded to disk, thus preserving flexibility for later on in the project.

Monitoring

Okay, so you've got a singer, mic, preamp, and A/D converter feeding a signal into SONAR where it will be recorded to hard disk. That's all good stuff, but there's something missing from that picture. How does the singer hear any tracks he is recording over, and, of course, his own voice at the same time? We'll assume the tracks are already recorded on hard disk and available in SONAR, but to be useful they need to make it out to the singer's ears, preferably in a way that will not be picked up by the vocal microphone. Also, the singer will need to hear himself in some reasonable balance with the tracks to turn in a decent performance. (Let's not even worry about optimal performances just yet.)

Unless you're brand new to recording with a computer, you can probably figure out what happens to get the already-recorded tracks out to where the



singer can hear them. SONAR plays the tracks back, they get converted from digital to analog at the D/A (digital-to-analog) converters, then the analog signals get fed to some kind of amplifier and finally to some form of speakers, probably a headphone amp and headphones, respectively, in the case of vocal recording. Most audio cards include D/A converters, and many also include headphone amplifiers, but those could just as easily be provided in some type of outboard form. In fact, unless you're recording your own vocals, you may need to go to an outboard headphone amplifier (or some other device that provides additional headphone outputs). This is because many sound cards that do feature headphone amps only provide a single headphone output, and you'll need at least one for the singer and one for the recording engineer.

What isn't quite as straightforward is how the singer's voice gets mixed with the tracks and out to those same headphones, while also being recorded to disk and synchronized with the recorded tracks. With SONAR, there are two basic options for how to accomplish this. Which is better will depend on your audio card and the singer's preferences:



SONAR features audio input monitoring that can be used with any sound card that will work with SONAR. When SONAR's input monitoring is used, the vocal signal comes in through the entire signal chain described in the section above, gets any effects and mixing added inside SONAR, then gets fed back through the signal chain described earlier in this section as part of the mix (or possibly some alternate monitor mix -- SONAR's buses and sends make it easy to create custom mixes) and through the vocalist's headphones. The good news with this approach is that you can use any effects you can use in SONAR as part of the overall chain that affects what the singer will hear without recording those effects to hard disk. Thus, the singer who will turn in a better performance by hearing herself in a very flattering light (i.e. whatever treatments you'd be likely to apply to her voice in the mix) can hear herself in all her processed glory. The bad news is that there is latency involved in getting the signal into, and back out of, SONAR. Any plug-in effects that have internal delays associated with them (e.g. many compressors) could add to the latency that comes from simply getting the signal in and out through the converters and audio card drivers. Also, keeping latency down as far as possible means using as small of data buffers on the audio card as possible, but using small buffers on the audio card means more work for your computer's CPU, which may mean hearing nasty artifacts if you have a processor that isn't capable of keeping up, or, alternately, raising the buffer size, which adds to the latency observed by the singer in the playback of her voice. This latency in the playback of the vocal can be disorienting to the singer if the latency time is long enough, and different singers will have different latency thresholds they can tolerate before their performance begins to suffer. This may mean some work on your part, for example to submix tracks to stereo so you can freeze softsynths and effects to leave as much of the CPU as possible for keeping input monitoring latency low. However, then it can also be harder to tailor the mix to the singer's preferences, for example, if she needs to hear more drums and less bass.

An alternative provided by many digital audio cards is onboard "zero latency" monitoring. Audio cards that provide this feature would generally provide some form of mixer applet that allows mixing input signals coming into the card with signals coming from computer-based applications, such as SONAR. The input signal is only going through the DSP (digital signal processing) on the card, not through SONAR. That is, the abbreviated signal chain for the vocal (or other audio input) is mic to preamp to A/D converter to audio card DSP to D/A converter to headphone amp to headphones. Of course, the vocal signal still goes into SONAR to be recorded, but latency inside SONAR is no longer a concern -- you can set it as high as you want, because the singer won't hear the latency. Thus CPU cycles are preserved for



SONAR's processing, rather than dealing with very small audio buffers. The audio card hardware takes care of letting the singer hear herself alongside the recorded tracks. The main downside to this approach is the singer will not be able to hear any processing of the input signal that would be done on the SONAR side since the voice she hears is captured prior to its being sent into SONAR for recording. Some singers will be perfectly okay with this. Others are only comfortable hearing their voice processed with effects like compression and reverb, similar to what might be used on their voice in a live show. Some, but not all, audio cards that provide this zero latency monitoring also provide a solution for this situation. Specifically, the need is for vocal processing effects that the singer can hear when monitoring her singing, but which will not be recorded into SONAR. There are two primary approaches to filling this need. One uses effects inserts, which can be used in conjunction with outboard effects modules, and which only affect the monitored signal. The other uses DSP-based effects that run on the audio card itself, where it is possible to have those effects inserted between where the vocal signal is sent off to SONAR and what the singer hears. The effects used in either of these scenarios may not be the ones that will appear in the final mix of the recording (though they could be in some cases -- again, it depends upon the capabilities of the audio card). However, they may be good enough for getting "in the ballpark", and that may, in turn, be good enough to make the singer comfortable.

Before we move off the subject of monitoring, there is one headphone-oriented monitoring tip worth mentioning. Sometimes no matter what you do, you just can't seem to get the right blend of the singer's voice and the background tracks in the headphones. Often this takes the form of the singer's wanting "more me", and you're happy to oblige him, but then, when you turn his voice up in the monitor mix, you can tell he's not hearing the instrumental tracks well enough because he's drifting increasingly off pitch. While adjusting the instrumental mix, for example to decrease or increase the amount of bass in the monitor mix, may be an option, perhaps you've already tried a number of combinations without success, and your time with the singer, and the singer's energy, are in short supply. Often the cure for this is asking the singer to try removing the headphone off one ear, so he is listening to the instrumental tracks, plus any monitor mix of his vocal, in one ear, and whatever is going on in the tracking room -- i.e. namely his voice singing in the room -- in the other ear. While this isn't a cure-all, the reason this technique can be effective, is because the request for "more me" often isn't so much about the volume of the vocal versus the tracks, but rather about the singer's not hearing the natural cues he uses to recognize his voice when singing. What is coming through the headphones in a close-miked vocal is missing most of the ambient cues we would get when we hear our own voices in a room. Effects can try to compensate for that, restoring some of what is lost to the close miking, but often simply hearing his voice in an actual room through one ear will give the singer enough of the cues he needs to get back to his comfort zone. Then you can probably reduce the volume of the vocal in the headphones, so the singer can actually hear the instrumental tracks he's trying to sing against again.

Capturing a Great Performance

Okay, so we've got an input signal chain to capture the singer's performance, and a monitoring chain to let the singer hear whatever is most useful to that singer. Our key goal at this stage is to optimize our chances for getting a stellar performance out of the singer, or at least as good a performance as the particular singer is capable of giving.

By the point the vocal session is ready to begin, it is too late to coach the singer in techniques for preparing for the session. Preparation time is over, and the singer is there with whatever level of vocal technique he or she has, in whatever state of sobriety (or not) he/she is in, with or without a cold, with or without "red light syndrome" (i.e. the tendency to freeze up when the red light signifying recording is in progress is lit), and so on. To quote one of my own song titles, you get what you get. The best you can do is accommodate whatever helps the singer get the best performance he or she can at that moment. This isn't a lesson in singer psychology, nor in herbal recipes for keeping vocal cords lubricated, so you're on your own for that stuff. What you can do, however, is try and make the recording process as transparent as possible.

To use an analogy, if you've ever tried taking photos of little children, you know that, as soon as they sense the camera is pointed at them, many (most?) of them tend to either freeze up, start hamming it up, or put on one of those pasted on smiles where you can just hear them saying, "cheese," whether or not they are actually saying it. While some children are more natural photographic subjects than others, and some can learn to "look natural on demand", generally the best results will come from candid shots, where they aren't seeing the camera at all or you've just

blended in with the environment to such a degree that they've forgotten you're there photographing them. Of course, the need to photograph a moving subject places extra demands on the photographer, so that often means taking more exposures, and possibly more time spent in cropping and other photo-editing activities later on.

Okay, so that's probably not a perfect analogy -- it's not like you're going to hide a mic in the singer's shower or anything. The idea, though, is to try and figure out how the singer will feel most comfortable, and do whatever you can to accommodate that. Note that this issue takes on special importance if the singer is you. Let's look at a few areas that come into play:

Does the singer prefer to perform a song all the way through, or break a song down into sections, tracking a little at a time, or even down into individual lines, tracking a line at a time?

If a singer sings a song (or a section of a song) several times through, is there a point at which she is most likely to hit her best take, and a point beyond that at which she hits the point of diminishing returns?

Can the singer jump into a section of a song and sing a line or two to replace one that was messed up in a previous take of the vocal and maintain a level of continuity and believability as if the vocal were recorded in a single performance?

Does the singer need a rest between two takes of a vocal part, or does he prefer to do several takes in a row to keep the energy flowing?

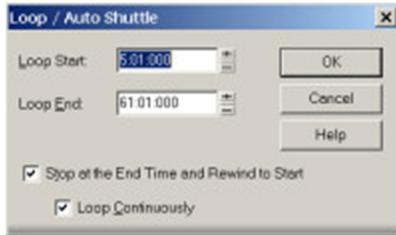
This list is not comprehensive, but the idea is that the singer's background and preferences will affect the tracking strategy that will achieve the best results. For example, some singers are not comfortable with the assembly line-like methods employed in tracking vocals in some studios, where you concentrate on an individual part, doing it as many times as needed until you have what you need, then move on to the next part, do that as many times as needed, and so on. Trying to force them into that style of tracking may result in stilted sounding recordings. On other hand, if you let those same singers perform the song straight through several times, keeping each take, even if it contains big bloopers in some areas, worrying about how you'll deal with repairing the problems later, you may get much more emotional performances, albeit with problems here and there. (This particular situation can be particularly common with singers who aren't very experienced, or just aren't very comfortable, in the recording studio, but are extremely dynamic live performers.)

The key is figuring out what tracking strategy will best suit the singer, and setting up SONAR to facilitate that strategy. The good news is SONAR can handle most any strategy you care to use. Here are a few common ones:

This first one is for the singer who is also playing recording engineer, takes a little while to switch gears from engineer mode to singer mode, and is very comfortable performing songs straight through as in a live performance, but not as comfortable in the recording studio. This profile fits me, and I'd guess it fits many readers. We may have our own home studios, and may be comfortable performing live, but we aren't experienced session singers. I've learned that my first few takes are more or less throwaways because I've been playing recording engineer up to that point, so have to switch mental and emotional gears. Not to mention that I probably haven't warmed up my voice, or maybe just haven't warmed up sufficiently, if I've been working on other aspects of the recording project prior to recording vocals. Takes 3, 4, and 5 will probably be where I start hitting my stride, and those may be all I need to feel confident I've got enough to yield a solid composite vocal. However, if I'm recording a new song, or one that has had recent rewrites, I may also be learning it and/or refining it in the process of recording the vocal, so additional takes may be necessary to get extra takes in any problem areas. However, after singing a song straight through five times, I can also reach the point of diminishing returns vocally, so takes 6 and beyond will often be more about emergency patch material than solid basis material.

The key in this scenario is to keep the time between recording engineer tasks (e.g. hitting SONAR's transport buttons like Record, Play, Stop, etc.; muting and unmuting tracks; changing loop points; etc.) as long as possible. This gives the singer time to get in the mood and up to peak performance. Then, while the singer is "on", it is time to perform as many takes as needed to ensure you've got solid raw material for creating a strong composite take

without interrupting the mood. At some point you'll either have enough or reach the point of diminishing returns, and can determine you've got enough done, and switch back to producer/recording engineer mode. We'll worry about "comping" (short for "compositing" -- the name for the editing that will get done to create a strong composite vocal from multiple takes) techniques later, but let's take a look at SONAR's facilities for supporting this tracking strategy.



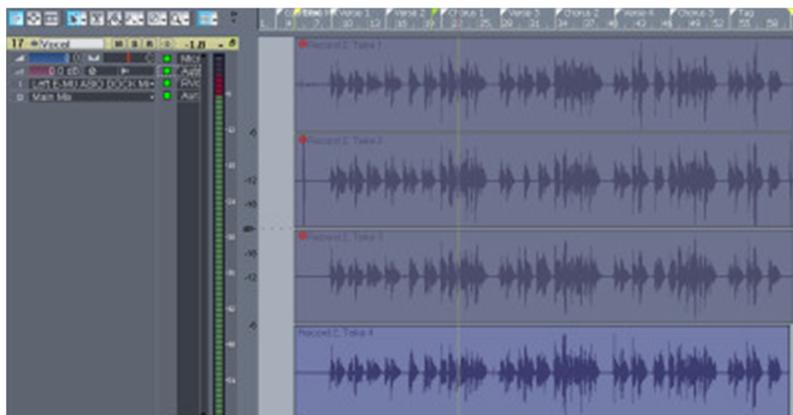
The first SONAR facilities that will help us here are the looping and auto shuttle features. You'll need to figure out how much pre-roll you need before the area you are recording to allow the singer (you?) to get into the performance (and to move from the computer to the microphone if you're recording yourself), then set the Loop Start point early enough to do that. You'll also need to allow enough room after the end of the singing to allow for any held out notes. If you need breathing room between takes, but don't want to stop the transport, you may also want to allow for that in setting the

Loop End point. While you can use SONAR's Loop/Auto Shuttle dialog box for setting the loop points, the easiest way is generally to select the portion of the project you want to loop over, either in the timeline for the project or by selecting a range of audio and/or MIDI clips, then use the Set Loop Points to Selection on SONAR's Large Transport or Loop/Auto Shuttle tool bar. I recommend setting the loop points on even measure boundaries, as it can sometimes be disconcerting to have looping occur mid-measure. You'll also want to set the Auto Shuttle options to "Stop at the End Time and Rewind to Start" and "Loop Continuously" to avoid having to play recording engineer in between takes. These are set automatically when you turn looping on in SONAR, so you shouldn't have to worry about them. The result is that, when you hit Record, SONAR will loop through the section you've set via the looping boundaries until you hit Stop, and you will only hear whatever you are monitoring when recording, not any takes recorded in previous passes through the loop since hitting Record.

What we've done thus far only sets transport behavior. It does not govern what happens with any recorded material. To tell SONAR what to do with the individual takes, you will use the Record Options. There are two main options for what happens with the audio record. One is to keep every take, and the other is to overwrite any previous takes with the current take. For our purposes, we will want to keep all takes, so we will select the "Sound on Sound (Blend)" option. This "blend" notion is somewhat of a misnomer in that the audio from the takes would, by default, be mixed together when you play back the track, but the audio takes are separate clips within SONAR, and do not get blended together in a single clip that is the equivalent of a bounced down mix of the takes. We will talk about how to control what actually gets played back a bit later. Related to keeping all takes is the decision where to store multiple takes, and there are also two options here. One is to store the takes in a single track, and the other is to store the takes in separate tracks. Prior to SONAR 4, dealing with comping operations when using a single track would be pretty difficult to manage due to user interface limitations, namely that all clips would be hidden behind one another if they overlapped, which they would in this situation. Thus, prior to SONAR 4, I would have recommended using the option to store takes in separate tracks, and you'd get a track for each take at the end of your recording operation. As of SONAR 4, though, SONAR provides a facility called Clip Layers that makes it easy to see multiple overlapping clips in a single track, and there are strong advantages for comping efficiency if you store the takes in a single track, so that is the way to go. There is one more option in the Record Options dialog box, and that is whether Auto-Punch is turned on or not. Auto-Punch sets a range for recording to happen, and that range may be shorter than your loop range if, for example, you want to have a lot of pre-roll and post-roll, but not take up all the extra audio file space to record during all of that. If you use this feature, you'll still need to make sure you are starting early enough and ending late enough not to cut off parts of the vocal, and that the Auto-Punch boundaries are within the loop boundaries. These days I tend to have enough hard disk space to not worry about using this feature -- I'll just record whatever extra audio may come from the pre-roll and post-roll, and slip edit it out later. (That "silent" audio can also be useful later on if you will use any kind of noise reduction process that requires analyzing a "silent" section -- i.e. a section of the audio with only background noise.)



Once you've got the Loop/Auto Shuttle Options and Record Options set, and have an audio track taking input from the port on your audio interface where the vocal mic is, and armed for recording, you're set to go. Just hit Record, run over to the mic, and start singing. Keep at it until you've done as many takes as you want, run back to the computer, and hit Stop. If at that point you hit Play, you'll hear all your takes playing at once, and you may only see one clip, despite hearing several takes. However, you can turn SONAR 4's "Show Clip Layers" option on by right clicking in the clip lane of the Track View, then use SONAR's clip muting capability to mute all but one of the takes for monitoring. We'll get to details of this later, but, once you've done this, what results may look something like the following:



In this particular case, four takes of the vocal were recorded. "Show Clip Layers" has been turned on to reveal all four takes. SONAR 4's clip-level muting has been employed so that only the fourth take will actually be played back for take auditioning purposes.

That was pretty simple, eh? The idea, after getting things set up, was just to let the singer, who is also the recording engineer in this scenario, sing. We haven't mentioned darkened rooms, lava lamps, herbal teas, and whatnot, but you can employ whatever you want to help ensure that the singer is in the mood and does the optimal job possible. From a recording engineer's perspective, though, all we are really trying to do is stay out of his way. Of course, the real work, which will be the main focus of this article, begins at this point. Before we get to that, though, lets talk about a few other tracking strategies that may be useful for singers with different preferences.

A minor variation of the above approach would be used in a case where the singer is not the recording engineer, and may prefer to rest a little between takes, but still prefers to sing the entire song in one go. Since there is no worry about the singer's needing to switch modes, like in the engineer/singer case, we don't need to worry about loop recording, so can leave looping turned off. The recording engineer can deal with the transport. Assuming you have SONAR's automatic rewind turned on, everything else in the setup can stay the same. However, between takes, the engineer will need to mute the clips for all earlier takes so they will not play back when the singer is recording the next take. This can easily be done with SONAR's Mute tool or simply by selecting any unmuted takes and hitting "Q" (the short cut to toggle the mute status of the clip) to mute them. If only one take is recorded each time, this simply involves clicking on the most recent take, then hitting "Q".

For a singer who only wants to record a section of a song at a time, whether that section is short (e.g. a single lyrical line) or longer (e.g. an entire verse or chorus), the techniques are similar. In the case of the singer who wants to record passes continuously, it is just like the first scenario given above, only the loop points will be set to a shorter region, and, once you are done with that section, then you will reset the loop points to another region and record some more. In the case of the singer who is not playing engineer and wants some rest between takes, the technique is exactly as in the non-looped case above, though needing some resetting of the recording start point between song sections. In both cases, you can use as much pre-roll and post-roll as you like. Independent of whether that pre-roll overlaps with another section of the song, SONAR's "Sound on Sound" feature will keep the individual clips from overlapping sections distinct, its clip layers will make it easy to see what is what, and you can do any slip editing needed to trim down overlapping regions, deal with crossfades, and so on later.

What about the singer who just wants to repair one word in a previously recorded phrase? Those of us who started recording with multi-track tape recorders may still have recurring nightmares about this scenario. The procedure went something like, "listen to the playback, with the singer singing along; at the precise right moment, punch in to start recording the word in question, then punch out as soon as it is over, hoping you didn't cut off the word before or the word after the small section you were fixing." When automated punch points were available in the tape deck you used, the scenario was only slightly better in that you could program the starting and ending points ahead of time, and, assuming you got them right, there was no worry about cutting off other words. You still had to worry if the singer's timing fit the punched in word cleanly enough, though, because a little overlap in either direction could make it sound funny.

In the SONAR world, you could use the Auto-Punch feature mentioned above to set up punch points, and use Overwrite instead of Sound On Sound mode to replace, rather than augment what is already there. However, if you're thinking about going this route, I have to ask, "why bother?" Why not just mute whatever it is you will be replacing, then record as little or as much as you want, just as you did in any of the scenarios above, and worry about slip editing things later on? You can still use Auto-Punch, if you like, to control how much before and after the punch point gets recorded, but I'd suggest setting boundaries that are longer than you are likely to actually need to provide room for error. While you'll still need to fit the punched-in word in the space available for it, even if you get too long a replacement word, there are editing tricks you can employ to shrink it down to size. If it's the right length, but goes to far in one direction or the other, you can just move it around. There are no worries about fitting something in an uncomfortably tight space in SONAR's non-destructive editing world.

Ah, but what if the reason the singer prefers this kind of no-return punching (or even tracking) mode is because she'll know what she wants when she hears it, and doesn't want anyone being confused by alternate takes once she knows she's gotten the right one? While you could use the Overwrite mode to enforce this sort of thing, I will again recommend you don't bother. Just use clip labels, and maybe even different colors, to label the supposed keeper clips as the main ones and the others as alternates, muting all the alternates. If you're as confident as the singer is in the singer's real-time judgement, you could always delete the alternates. However, I'd recommend you keep those around, just for safety's sake, in case someone discovers a nasty problem later. You can always use those alternate takes as sources of "bandage material", to fix up any problem spots in the supposed keeper takes that may have gone undetected in the heat of the moment.

I've been talking primarily about recording lead vocals to this point. If you're recording non-group background vocals -- i.e. with only one singer singing a part into a single mic, where there is no bleed between multiple singers and mics -- there is absolutely no difference in tracking lead and background vocals, at least from the perspective of SONAR tracking techniques. (Depending on the musical style, song, and singers, there will often be differences in mic and singing techniques. There will also be potential differences in handling the recorded takes later on.) If you are tracking multiple background vocalists singing into a single mic (or multiple mics where there is significant bleed between singers and mics), the actual tracking techniques will still remain the same. However, in this case, greater attention may need to be paid to quality-oriented issues such as pitch control and subjective issues such as the blend of voices as you will have much less possibility for fixing that sort of thing later in the editing and mixing stages. (For example, while you can easily tune the recorded vocal of a single singer who is slightly off pitch, you can only do that with multiple singers singing a chord if they are all deviating from their desired pitches by the same amount.) Thus, it is quite possible you may need more takes than you would when tracking single singers, given an equivalent caliber of singers. (On the other hand, talented, experienced background singers may well find it easier to achieve tighter, more pitch-perfect results singing together in a group than singing their individual parts in isolation against recorded tracks where the other parts may not even have been tracked yet.) One specific thing I would recommend, if it is feasible for your situation, is postponing tracking background vocals until after you've got your final, comped lead vocal. The idea is that the background vocalists then can sing against the real lead vocal, not some slight-to-extreme variation of it, and this will save you editing time later on, assuming reasonably talented, experienced background singers, anyway.

The bottom line at this stage of the vocal recording process is to make sure you have whatever raw materials you may need to come up with whatever level of polished end results you want to achieve. That is always a judgement call, and may take some experience to consistently make such that you achieve your desired results. As a general rule, better singers tend to require less raw material to get any given level of polished results, while less experienced

and/or less talented singers may need more takes to achieve even much less-polished results. In general, it is better to err on the side of recording more than you need, until you gain confidence in your ability to judge when you've got enough. However, there are tradeoffs, such as tiring the singer out more by doing more takes, frustrating the singer who may think he's already gotten enough good stuff down, and -- the one that will affect you personally as we move past tracking -- increasing the amount of material you'll have to review and handle during the post-tracking phases of production.

From Raw to Polished

In the perfect world, your tracking would have gotten you a single perfect take of each vocal part, and you'd now be ready to add any desired processing and work on your mix. I don't know about you, but I've worked with some really good singers, some pretty good singers, some okay singers, and some pretty lousy singers, and I can't recall this perfect world scenario's having occurred even once. Now, there were a few times where the pressures of time didn't give me any choice but to act as if I had the perfect vocal take in front of me and proceed straight to processing and mixing. Given the benefit of extra time, though, I'd have preferred to take that time to work on the raw materials I had in front of me to improve upon them, rather than making whatever compromises were necessary to meet my timing constraints.

The thing is, you can have a superb singer, who delivers extremely moving live performances, with good-to-great pitch control, excellent phrasing, good dynamic control, and lots more, where you just hear that singer's performance, and think, "that was absolutely perfect, and I can't imagine improving upon it!" Then you can listen back to that same performance on tape or hard disk a bit later, when you don't have the benefit of the visuals, the emotion of the moment, and so on, and you do have benefit of a Rewind and Play button to listen repeatedly, and think, "this could be better!" Or sometimes, you might even get to, "what was I thinking when I thought it was a perfect performance!" Something that goes by you once and sounds fine, or even great, may wear on you with repeated plays. Remember, we are talking here about creating a recording that will hopefully be played over and over, preferably by many, many people, not a one-time performance that benefits from a concert atmosphere and the visual side of the singer's performance. Not to mention that many singers just aren't anywhere near as comfortable performing in a recording studio as they are onstage, but may still want to hear a recorded performance that comes as close as possible to the one they can deliver, or at least think they can deliver, live. Then technical and commercial limitations can enter the picture, such as whether the mics you have available for the recording the singer are good matches to the singer's voice, the neighbor's dog's barking in the middle of an otherwise excellent take (what, you didn't soundproof your non-commercial home studio?!?!), and so on.

In the remainder of this article, we'll be looking at techniques that can be used to deal with many different issues that detract from your having the perfect take you know the singer is capable of, at least in your optimistic view of her potential. To those who might be inclined to think some, or all, of this is "cheating", let me remind you that humans are fallible. The little, or big, mistakes we make during a live performance that may go unnoticed in, or even help give extra character to, that live performance, can easily become annoyances when they are repeated over and over on our CD player or (thinking extremely optimistically) the local Clear Channel outlet. We're trying to create a level of performance that does not exist in reality, because every live performance is unique, but which must exist if it is to stand up to repeated listening and the increased scrutiny that implies. Also, the available of computer-based editing and processing tools has become so widespread now that the bar has been raised for things like pitch control in recorded vocal performances. This can make deviations that would have been perfectly acceptable a few decades ago stand out like a sore thumb when played alongside the modern, pitch-corrected performances that seem to be the norm in today's production-oriented world.

So, how do we get from raw to whatever level of polish we desire? There is no one right way. Much will depend on your particular preferences in terms of editing and processing workflow. We will be breaking things down into separate techniques that can generally be used in most any order to deal with whatever issues you are trying to resolve. However, we've got to start somewhere, so I thought I'd begin by outlining my own "typical" (if there is any such thing) workflow. I will then use that outline as a guide for sequencing the detailed techniques I'll be discussing in the rest of the article.

Here is my basic flow:

1. **Lead vocal tracking:** My general process is to record anywhere from 3 to approximately 9 takes, with the typical number being in the 3-7 range for a recording intended to be a polished master, and 3 for a song intended as a demo. If there are any particular problem sections, I may also do additional takes of only those sections. We've already discussed all we're going to cover on this topic earlier in the article, but I wanted to clearly indicate where it exists in my overall vocal tracking, editing, and processing workflow.
2. **Cutting the vocal up into phrase-length clips:** I generally try to break the vocal down into phrases, generally separated by breaths, as they tend to be natural units for comping.
3. **Adding rough processing to the vocal:** The idea here is to use an approximation of how the vocal will later be processed such that any comping decisions which might be affected by processing, especially compression and limiting, can be made in light of similar processing. I generally also include automatic pitch correction in this processing so minor pitchiness will not influence my comping decisions -- I'm going for emotion, not technical perfection; we can address correctable technical issues ourselves.
4. **Auditioning and comping:** This is where I pick all the various bits I'll use and throw away. Because I've already broken things down into phrases, it is primarily a decision process, rather than a technical process. However, there may be some technical tasks involved, too, in the case of problem phrases, where it becomes necessary to break things down even further than the phrase level. When that happens some additional cutting things up, and possibly editing (see below), may be needed.
5. **Tuning:** While I've used pitch correction in automatic mode for auditioning to this point, I haven't actually done any permanent pitch correction. I find automatic pitch correction generally doesn't yield as good of results as manual pitch correction, and I don't want to be manually correcting the pitch on a bunch of stuff I'll be throwing away (and there is usually more that gets thrown away than kept).
6. **Editing:** This can involve anything from manipulating clip fade-ins and fade-outs to dealing with crossfades for overlapping clips, shifting clips around to deal with any timing/rhythmic problems, stretching or contracting syllables and/or phrases to deal with timing problems or editing challenges, and more. The basic elements I always do are easy and quick, but more complex issues can get into what I like to call "microsurgery", and can even get down as low as borrowing consonants from one word to graft onto another word to fix diction problems.
7. **Background vocal (and lead vocal double) tracking:** While it is not always possible, I strongly prefer to defer background vocal tracking until there is a final (i.e. comped, tuned, and edited) lead vocal to track against. Doing otherwise typically means more work in the subsequent stages of this process since the lead vocal may be a moving target, for example if there are phrasing problems. I am also lumping any lead vocal doubles in with background vocals, as I generally treat them similarly with respect to work processes.
8. **Background vocal comping/tuning/editing/processing:** I'm lumping these into one step here to simplify what is essentially a repeat of what has already been described above for the lead vocal. However, because the background vocals will generally be less prominent than the lead vocals in the mix, and there are generally many more of them, I will tend to take as many shortcuts as I can get away with to speed up the process.
9. **Submixing background vocals:** This step may be optional for people with really fast computers, but many of us will need to do this, and that means getting things like individual part processing, pan positions, and balances set in stone prior to the mix.
10. **Final processing, fader automation, and mixdown:** If the effects on the lead vocal and background vocal submix haven't already been finalized, this is the point to get that done, be it as a separate stage prior to the final mix or as part of the mixing process. I tend to like to get the processing set fairly early in the mixing process so it can be taken into account when dealing with other, less important, components of the mix.

The Guinea Pig

For purposes of illustrating this article with screen shots to demonstrate the techniques, and audio examples to demonstrate the results, I needed to pick a project. While most of my recent work has involved producing another vocalist, who is a much better singer than I am, I didn't want to embarrass anyone else by presenting "before" results which would never see the light of day on an end recording. This meant choosing one of my own projects. (Don't worry, I don't get easily embarrassed.) Unfortunately, my recent focus on producing someone else didn't leave me with a very big selection of candidate projects. In fact, most of the projects I had lying around either were already beyond the "before" stages I'd need for early examples, or had sufficiently far to go that it would take me quite awhile just to get other aspects of the projects together to some presentable degree. Thus, the guinea pig project is ... (drum roll please) ...

Awhile back, when I was working on a review of the [JMT Orchestrator](#) arrangement generator, I'd tried to very quickly put together quick generated arrangements for a bunch of old songs which I could never quite bring myself to demo, but which still had enough redeeming qualities that I figured I'd like some kind of recording to document. I wanted to try various styles, and remembered I'd had one very old song (we're talking 25 years old -- vintage 1980!) that I thought would suit a bluegrass treatment. The song is called "Baby", and I'd originally just put together JMT Orchestrator tracks, with a variety of stylistic changes to vary the arrangement over time, then added a synthesized lead vocal with [Zero-G's MIRIAM Virtual Vocalist](#).

I might have left it at that, since the song was documented and I didn't really see any commercial potential in the song. However, a few people who heard that recording liked it more than I'd expected, and mentioned they'd be interested in hearing how it would sound with a real singer. As a result, later on, when I was working on a review of the [E-MU 1820M](#) audio interface, trying out the built in mic preamps, I added this song onto a vocal tracking session where I knocked off 3-4 songs in the space of an hour or so. In particular, I'd tracked 4 raw takes of the lead vocal for this one in one continuous, looped recording pass. Since that time, I really didn't have any significant pockets of time to work on a low priority song. Thus, it has pretty much just languished on the "to be finished" project pile, although I did at least go in and replace some of the bluegrass instrument sounds I'd been using previously with some new, high quality sounds from the East West Quantum Leap RA software instrument.

The vocal tracking chain for the song was a CAD Equitek E-200 mic (a condenser that cost about \$500 back in early 1996) to the E-MU 1820M's TFFPro preamps and high quality A/D converters into SONAR. There is no special acoustic treatment in my small home studio, which is a small, squarish room with standard 8-foot ceilings. My tracking position is roughly in the middle of the room, with my mic facing an open door (thus minimizing reflections from behind my singing position). My computer, including its fan and hard disk noises, is positioned at roughly the 90 degree point to the front of the mic, but on the floor under a desk and probably about 4-5 feet from the base of the mic. My U-shaped desk may buffer a little of the computer's noise from getting to the mic, but there is still plenty of that noise that makes it through. The mic was used in its cardioid pattern for maximum off-axis sound rejection, and I did use a pop filter roughly 3-4 inches from the mic, with my vocal position being another 4-8 inches out, depending on the part of the song.

As for the caliber of the singer, you can make our own judgments, but it's worth mentioning a bit of my own self-assessment since the clips I'll be using are short. In general, I've been singing every since I've been talking and love to sing. However, I am most comfortable singing while also playing the piano, and am decidedly uncomfortable standing in front of a microphone and trying to capture "the perfect take". I think this may also have been the first song I tracked in the session, and I probably hadn't warmed up at all prior to starting singing the song. When I start worrying about pitch control and various other technical aspects of vocal recording, I tend to exacerbate any problems I might otherwise have (e.g. worrying too much about staying in tune tends to make me sing all the more out of tune). Thus, I have learned to do my best to not worry at all about the technical side of the performance, and just do my best to get into the lyric of the song as an actor might, going for believability and emotion over vocal technique. When I am successful in doing this, the results can be reasonably good, but I'm not terribly consistent in doing that when standing in front of a microphone (I'm much more consistent while singing and playing at the same time). Thus, the raw recorded vocals tend to be full of flaws, mixed in among promising moments.

For anyone who is interested in hearing what some of my relatively polished results can sound like, I will refer you to my music page on SoundClick.com (http://www.soundclick.com/bands/4/rickpaul_music.htm). And now get ready to cringe, because here is the first "before" clip from our vocal editing guinea pig, "Baby" (words and music by Rick Paul, © 1980 R. Paul, All Rights Reserved):

["Baby" - Raw first verse vocal from take #4](#)

I've made the clip the soloed vocal so you can more easily hear any flaws in the vocal or recording, with no distraction from any underlying instrumental tracks. However, it can be helpful to have some idea of context, so here is the same raw vocal take mixed, with no vocal effects, along with the JMT Orchestrator-generated temp tracks used for tracking the vocal:

["Baby" - Raw first verse vocal \(take #4\) plus temp tracks](#)

With that "before" picture in view, let's move ahead with trying to take the raw materials and fashion a more compelling recorded performance.

Preparing to Comp: "Comp-able" Units

The setup in the "Guinea Pig" section above was partly intended to set the perspective that our audio examples were recorded in a fairly modest recording studio. However, it was also designed to provide enough background information on the particular singer whose vocals we'd be editing to give some clues on picking an editing approach. In particular, I'd noted the raw recorded vocals tend to be "full of flaws, mixed in among promising moments." You haven't heard all the raw takes, so you'll just have to trust me on that characterization here. In general, though, one of the first things you will need to do after tracking raw vocals is characterize the raw materials with which you'll be working, so that you may decide on a comping strategy.

There are two primary strategies for comping vocals. One is to assume we will mostly work with one "best" take and only comp when there are problems in that take. The other is to assume there is really no such thing as one best take, and we'll need to comp everything, treating all or most of the takes as being relatively equal with respect to their potential for being the source of raw components for the composite track. In between those two extremes there is a position where we might expect two or three of the takes are likely to be better than the rest. These may be the primary sources for raw materials for the composite track, but there may still be a need to look outside those tracks in some cases. When it comes to preparations for comping operations, though, this intermediate strategy has more in common with the "no best take" strategy than it does with the "one best take" strategy, so we will focus on the more extreme positions for the moment.

The "one best take" strategy will tend to suggest you have one of two types of vocalists. One of these would be a fairly predictable vocalist, who always hits her stride on a particular take then goes downhill from there. The other would be a really good, and consistent, vocalist where every take is pretty much magic, so it doesn't so much matter which take you use as a starting point, though there may be a few problems here and there. There is one other scenario where this strategy make sense, and that is where you don't have either of these two types of vocalists, but you do have a very short deadline, and have to favor keeping the amount of comping time short over going for perfection.

In the "one best take" strategy, the main preparatory step will be to audition the various takes to figure out which one is the best starting point. When you get to actual comping, you will mainly be listening serially through the performance, trying to identify trouble spots, and finding alternate takes to substitute in those spots. You won't pay much attention to the alternate takes unless there are problems in the "best" take, though. As such, there is really no need to break the individual takes down into smaller units ahead of comping.

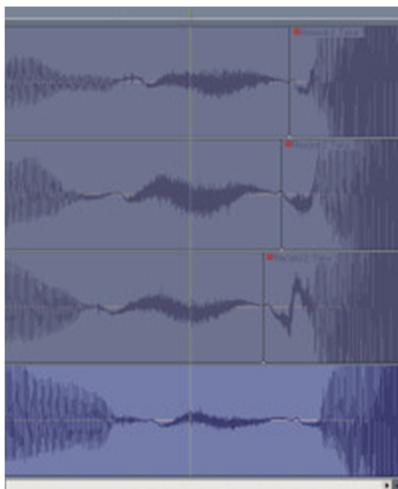
By contrast, with the "no best take" strategy, or the related "multiple good takes" strategy, we can assume we'll be doing a lot of detailed auditioning of candidates for any given part of the performance. Thus, it makes sense to break

the performance down into what I'll call "comp-able" units. What is a "comp-able" unit? Well, it's something I totally made up, but the idea is to find portions of a performance which make it easy to mix and match multiple takes of those portions with multiple takes of other portions of the performance. For vocals, there tends to be a very convenient unit, which I'll call a "phrase", which is bounded by the vocalist's breathing. While this isn't necessarily the same as a lyrical phrase, it often is, and this can help for keeping a solid flow within each phrase. However, the key reason for picking breath-separated phrases is because it generally provides a convenient point to chop things up, where there is typically not much happening in the vocal, and there would be a natural break in that vocal while the singer breathes. This can mean transitions from one take to another can be pretty seamless, especially if you compare it to something like trying to make a transition in the middle of a word or even between tightly-spaced words within a phrase.

We'll be using SONAR 4's Mute tool later on to help us audition alternate takes of each phrase. For this "no best take" comping strategy, we will be using the "mute entire clips" mode of the Mute tool, so, to prepare for comping. We will cut each vocal take up into phrases, using SONAR's Cut tool, which can be invoked by hitting "C" on the keyboard or simply clicking on the Cut tool icon (see the graphic at the right of this paragraph) to turn the tool on. First, though, if you haven't already turned SONAR's "Show Clip Layers" attribute on for the track you'll be comping, now is the time to do that. Simply right click in the clip pane area of the track in question, and enable "Show Clip Layers" in the pop-up menu. You should be seeing multiple clips, one representing each of your takes, within the single track you'll be comping.



Where do we cut? As indicated above, we will be cutting the vocal down into phrases delineated by the singer's breathing. Thus, we are looking for points at which the singer took a breath. There are two ways to find them. One is to solo the vocal and just listen, hitting Ctrl-Spacebar to stop the SONAR transport at the point where you hear a break in the phrase, then use visual cues to find the breath in the audio wave. The other is to simply scan through the audio waveforms visually to locate breaths, which tend to be extremely easily recognizable once you know what to look for in the waveform. Even if you're using audio cues, to save time getting into the general vicinity of a breath, you will ultimately need to use visual cues to find the best point for cutting, so let's take a look.



First, you may find that the singer's breaths were too quiet to actually show much of anything on the visual waveform. If so, you can use SONAR's capability of audio scaling to increase the size of the waveform within the audio clips. Simply hold down the Ctrl key while clicking the zoom button that would normally zoom in vertically in SONAR's clip pane. The view you are looking to achieve is one where you can easily see starting and stopping points in the singing and the singer's breaths, but where you are not zoomed in so far as to make it tough to distinguish between foreground audio, be it desirable or not, and background noise. The screen capture at left shows a portion of four takes' worth of audio waveforms where one vocal phrase ends and another begins, with a breath in the middle. Audio scaling has been employed to give the breath enough prominence to be easily recognizable against the background noise.

So where do you cut? You can cut either before or after the breath, or even on both sides of it in cases where you may end up having some overlap issues where it might be useful to consider the breath as a separate phrase for comping (e.g. to use a shorter breath in a tight space). Note that you generally will not want to cut the breath out altogether, because vocals without breathing sound pretty unnatural, and extremely clinical. The breathing is just one of the vocal cues that helps us feel the singer's emotion. As to whether to cut before or after the breath, the answer is, "it depends." Where can you make the cleanest cut, to avoid chopping off part of a word? If the breath just flows out of one end or the other, then cut at the other end. If you could go either way, then where can you make the cleanest alignment of cuts between takes? In the example at right, take #4 has the last note in the first phrase being held out quite a bit longer than that same note in the other takes. If I were to cut before the breath, the cut point in take #4 would be overlapping with the breath in all the other takes. While that isn't necessarily the end of the world (you can usually get away with truncating a breath and fading it in or out to avoid an abrupt transition), I can cut after the breath and get much better aligned cut points. There will be virtually no overlap between the end of the breath for any of the takes and the start

of the next syllable in any other takes. In fact, the only overlap that will exist at all will be between the end of the breath in take #4 and the very start of the syllable in take #3. That would represent an easy crossfade if it turned out we'd be putting the end of the first phrase from take #4 together with the start of the second phrase from take #3 in the composite vocal. Thus, I will choose to cut after the breath (the graphic shows the cut for the first three takes only).

One additional note on cutting is that it is important to cut the waveform at zero crossings to avoid clicks in the audio. SONAR will handle this automatically if you leave the "Snap to Audio Zero Crossings" flag enabled (it is enabled by default) in SONAR's Snap to Grid dialog box. The bottom line is you want to try and cut in a silent area of the take. Due to background noise, there probably are no true silent areas, so you will be cutting in areas where there is only background noise.

There is one additional consideration in this breaking down of takes into phrases. That is whether you will do this a little at a time, then comp that section, then move on to the next section, or you'll make a pass through the whole song to cut things up, then make a comping pass after that. I generally prefer to go the latter route because the slicing things up pass becomes a pretty mechanical pass, and, once I get in the swing of zooming and cutting, and possibly deleting extraneous (i.e. "silent" or "just noise") clips, I can go pretty quickly through a song. Whereas the comping pass is an extremely aesthetic pass, where I am trying to make musical, albeit with some technical considerations, decisions, and I don't want to keep switching gears between the aesthetic and the mechanical. However, SONAR's performance can sometimes bog down when you get a lot of small clips going (at least on my Athon XP 1600+ system), so there are tradeoffs. It comes down to personal preferences.

Preparing to Comp: Minimizing Distractions

We're almost ready to start comping, but, remembering that comping is primarily about making aesthetic decisions, there is one additional preparatory step I'd recommend making first. That step is trying to minimize any technical issues that distract you from concentrating on the aesthetic considerations upon which you will base your comping choices. Let me give a few examples:

- **Minor pitchiness:** If a singer is way out of tune on a phrase within a take, odds are I won't be selecting that take when I comp the phrase. However, if there is some minor pitchiness, to the degree that it sounds a bit off when listening to the raw vocal, but within the realm of what can easily be handled by pitch correction software and techniques, I don't want that pitchiness distracting me from what might otherwise be a great take. After all, I can fix it in the final vocal, so it effectively won't be there and shouldn't be considered a negative when considering the take.
- **Minor dynamic inconsistencies:** Sometimes a singer will just mess up and sing a word or syllable way too loudly or softly. This yields an uneven vocal where it isn't really possible to fix the problem via fader riding or dynamic processing (e.g. compression and limiting) because things are so far out that fixing the problem causes other side effects (e.g. raising the background noise too high). Other times a vocal line may be a bit inconsistent in dynamics, but is within the range of what just running the voice through a compressor and limiter could cure. Still, those minor inconsistencies might distract from the performance if just listening to the raw vocal.
- **Predicting the results of vocal effects:** Rarely will a raw vocal make it into the mix without use of some form of effects. Those effects may include compression and limiting, equalization, de-essing, aural exciting, delay, reverb, and much more. In some cases, those effects may affect some basic vocal characteristics, such as timbre, and the differences might be significant enough to cause you to choose a different take if you were choosing between the raw takes, as opposed to choosing between the processed takes. Where this is the case, it is advantageous to listen to the something processed similarly to how the final vocal is likely to be processed, so you can consider your decisions in a context that is closer to where your final mix will go.

We'll talk about the vocal processing chain, and various types of effects that are relatively commonly found in it, later in the article. For now, though, I will simply give you an idea of what I will typically configure at this point, then discuss one particular effect that will not be in my final vocal chain which can be invaluable at this point in the vocal production process.

When it comes to configuring my own vocal effects chain for comping purposes, I am mainly considering effects that will affect the timbre and dynamics of the vocal because vocal dynamics and timbre can vary from take to take, and even within a single phrase of a single take. It is possible that processing may exaggerate some timbral differences while hiding others. Thus, if there is a strong likelihood I'll be using such an effect in my end mix, I'd prefer to have it in the chain for auditioning effects. I don't worry so much about ambience-related effects, such as delay and echo.

One particular issue I face in my recording setup is that the one microphone I own that is a reasonable vocal recording mic is badly matched to my voice. It tends to exaggerate some of the worst aspects of my voice, and sounds nothing like what I am hearing in my head, nor like what others hear in a live performance scenario. (Unfortunately, I didn't realize this until it was already too late to return it and get another, more suitable, model.) After a few years of trying, unsuccessfully, to remedy this with EQ and other types of processing, I came across Antares' Microphone Modeler, which attempts to subtract the characteristics of your mic and add on the characteristics of any mic you choose from within the models whose characteristics have been modeled for use with the software. Because mic characteristics tend to affect timbre in different ways depending on pitch, volume, and more, this is something I always try and get set as I'll want it early, before adding other effects and definitely before making any comping decisions. In this particular case, I've changed my CAD E-200 into a "virtual AKG C12A" -- anyone wishing they could do that in the non-virtual world?

My voice tends to be slightly sibilant (i.e. prone to exaggerated, or hissing, "s" sounds, with some exaggeration in other related consonant sounds). While it can be fine on its own for recording, I also frequently use a plug-in that adds harmonic excitement to my voice (to help it cut through a mix better without adding EQ in a range that can make my voice too strident), and that plug-in tends to enhance sibilance to a degree. Sometimes that is just fine, and other times it puts things over the edge. This particular test case is one of the latter, so I put a de-esser (Waves RDeEsser in this case) after the Microphone Modeler, and placed the harmonic exciter (DSP/FX Aural Activator here) later in the chain.

I don't always use EQ on my voice, but, with this particular set of tracks, it seemed to help toward fitting my voice in the rough mix, so I added that (PSP MasterQ in this case) after the de-esser, then followed the EQ with the harmonic exciter. Last up was a compressor/limiter/gate (Waves RVox) here, to even the dynamics out and ratchet things up in the "in your face" department.

Again, we'll be talking more about effects later, and I should emphasize that the chain I am using here has no special magic to it. I often switch the order of some of the effects, for example whether EQ comes before or after the compressor (if EQ is used at all), and even what is in the chain in general. While I often use the same plug-ins with my voice, I might use quite different ones if working with someone else's voice. Most of the time some degree of experimentation is involved, and the real bottom line is how it sounds, especially in the context of the other instruments in the mix, which is a rough mix at this point. You can hear the changes auditioning the same phrase with effects makes here:

["Baby" - Raw take #4 vocal with processing](#)

One effect that will not be in my typical vocal chain come mixdown time, but which I find invaluable to include for auditioning takes at the comping stage, is an automatic pitch corrector. I will use Antares' AutoTune 4 here and later in our vocal processing work. At this stage, we are not actually interested in tuning the vocals so much as we are interested in minimizing the distraction from easily fixable tuning problems and making less easily fixable tuning problems more noticeable. The idea is that an easily fixable problem should not be allowed to distract from choosing what is otherwise the best take (i.e. emotionally, and with respect to other aspects of the performance). On the other hand, if a tuning problem may not be so easily fixable, it would be nice to have it called dramatically to our attention

so we can either avoid using that take or experiment with what will be involved in fixing it, just in case our decision for a particular part of a performance comes down to a number of contenders among which none is an obvious choice, where the decision may hinge upon whether we can get around a more severe tuning problem in an otherwise preferable take.

For this application, we will use AutoTune's automatic mode. Since I know this particular song is in the key of B-flat, and all notes within the song are within the B flat major scale, I will set the key and scale to be used as a reference B-flat and Major, respectively. If you didn't know the key of the song, or whether the melody of the song stuck to scale notes within that key, a good default position is to set a chromatic scale in any key. Whatever is set here, AutoTune checks the pitch of the input vocal against notes in the reference scale (more technically, it is checking frequencies in the input vocal against reference frequencies corresponding to notes in the reference scale), and shifts the pitch up or down to the nearest scale frequency after accounting for some timing parameters such as how quickly to retune a note after discovering it is off-key (at least according to the tracking and tuning criteria set), how choosy to be about divergence from reference pitches, and more.



In the default case of a chromatic scale, this means that if a note is off-key by less than a half of a semi-tone, AutoTune will tune the note to what it should have been. However, if it is off-key by more than half semi-tone, AutoTune will tune it to the wrong note, because it is actually closer to that wrong note than the desired note. Thus, if you are familiar with the melody of the song in question, minor pitch variations go more or less unnoticed while larger pitch variations get exacerbated and turn into wrong notes. When using a reference scale with coarser granularity than a chromatic scale, some notes of the scale can be more off-key than others before they get tuned to the wrong note, but the principle is the same. In particular, if you have automatic tuning on, and don't notice something strange about the part, odds are you can do a good job of tuning the part and don't have to worry about pitchiness when choosing parts to include in the composite vocal. On the other hand, if you get obvious wrong notes, then odds are you'll have a harder time correcting the pitchiness in the part. At that point, you can eliminate that take of the phrase from consideration in the comp. Alternately, if you feel the other options aren't great, either, you can try changing the scale to only include notes in the phrase, or using AutoTune in graphic mode, where you can specify exactly how the melody is supposed to go (we'll talk about this mode more below when we get to actually tuning the vocals), to see if the pitchiness problem can be resolved before making your decision.

This isn't an AutoTune tutorial, so we won't go into all the parameters of AutoTune's automatic mode here. The main other things you may want to set are the Input Type (e.g. Soprano Voice, Alto/Tenor Voice, Low Male Voice) and whether "Improved Targeting" (more on this below when we get to tuning, but this is a setting you'll need to experiment with to see whether having it on or off yields best results) is turned on or off. Again, we aren't actually tuning the vocal at this point, but only using AutoTune to help us distinguish between easily correctable pitch problems and pitch problems that may indeed be problems that should eliminate a take from consideration for the phrase under review. You can get an idea of what running AutoTune in automatic mode does to the picture by listening to the following clip of the same vocal part we've been using to this point:

["Baby" - Raw take #4 vocal with processing and automatic AutoTune to B-flat major scale](#)

One thing you might notice here is this performance just got a lot shakier-sounding on the pitch front. (Perhaps it is not quite as obvious without knowing the actual melody of the song, but I think some of the issues will be obvious even without that.) This is because my performance on this take was pretty bad in the pitch control department. (Simon Cowell and Randy Jackson would crucify me -- even Paula Abdul might admit my pitch control could have been better; she's always so nice!) In fact, it was bad enough that AutoTune exaggerated the problems, and made them much more noticeable, so I can be listening with a more aesthetic ear and a less technical ear when I'm considering the takes.

I might add that I put AutoTune between Microphone Modeler and RDeEsser in the vocal chain for auditioning the clips. I wanted it after Microphone Modeler because Mic Modeler is a mono-only effect, and all the rest of the effects in my chain can handle stereo. I also wanted it somewhere prior to the compressor because the compressor

would emphasize any tuning-related glitches or artifacts that might occur, which might be an important consideration down the line.

Basic Comping

Depending upon your perspective, we've now either reached the fun part or at the point where will we be tortured. I'm inclined toward the latter view, especially if there are a lot of takes (e.g. anything more than three). There are two main issues that make comping painful. One is the scenario where you've got some number of takes, but none of them are clear standouts, and each of them has at least one issue that might make you eliminate it had you had a better slate of choices. The other is where all the takes sound pretty good, albeit with subtle differences between them, and it's hard to clearly eliminate any of them directly. In either case, you may have to figure out some other considerations, such as how they subtly effect the overall dynamic of the section when used with the leading choices for earlier and later phrases, where there also may be no clear winner. This multiplies the number of choices you have to consider due to having to combine earlier and/or later phrase choices with the choices for the phrase you're actually comping. It can get to the point where you feel like you're about to go crazy from too many non-optimal choices.

On the other hand, this is the point at which you can take a set of "pretty good" or "good" takes of a vocal performance, and combine them in a way that makes them into a "great" performance. Or it can equally let you take a bunch of not-so-good performances, but with various redeeming characteristics spread throughout the various performances, and turn them into a worthwhile performance. The bottom line here is no one will ever hear the raw takes, or know what you went through in turning those raw takes into a polished result. It is also the key point in the process for determining whether the result people will hear will be bad, ho-hum, pretty good, good, or great.

What are we looking for here? This will come down to your personal preference as a producer. That's right, I said, "producer", rather than "engineer", because comping is mostly about making musical and aesthetic choices, even though there are some technical procedures and decisions involved. Here are some key things I tend to look for:

- **Emotion and believability:** The combination of these two highly subjective traits is, in my opinion the **biggie**. I want to believe the singer when she sings the lyric, and I want her conveyance of that lyric to affect me emotionally. Of course, even a great actor can't save a lame script, and the same is true here if the song is bad. However, given a particular song, the difference between a good performance and a great performance often comes down to how much we believe the singer.
- **Mistakes and unresolvable technical problems:** I actually often eliminate gross problems earlier, when cutting things down into clips. There is no use in having to review five takes of a phrase if two of those messed up the lyrics. (Of course, if the three remaining takes also have big problems, then we might still want those extra two takes to be around if there are parts of them we can use to cut things up more finely.) However, lesser problems can help in eliminating subpar takes at this stage.
- **Consistency:** Sometimes a singer will start out singing meekly, build confidence over the course of a few takes, then start tailing off as his energy wanes. The vocal tone may differ significantly between takes, and we ultimately want the composite vocal to sound like it is sung by one singer in a single performance, independent of how much surgery may have been involved in piecing it together.
- **Dynamic flow:** This relates somewhat to believability and emotion, but the idea is to have a performance that builds dynamically in relationship to the dynamics of the lyric and melody. If a singer comes out of the gate going great guns in verse one, and that seems super for this song, but then tails off in verse two when the lyric calls for building intensity, the dynamic of the song will be off if you use those high energy takes for verse one. Thus, we are looking to make the dynamic flow of the performance parallel the dynamic flow of the song itself.

- **Uniqueness:** Many lesser experienced singers, and even some singers who are very experienced, especially as backup singers, have it all together when it comes to breath control, pitch control, dynamic control, mic technique, and various other technical aspects of singing. They sound great no matter what they sing, but, if you heard them singing alongside several other singers of similar caliber, you might have a hard time distinguishing one singer from the other. The reason is because they haven't yet found their own voice or uniqueness -- i.e. those special traits that let you recognize their voice within one or two notes, like you can do with almost every singer who has ever made it to superstar status. I have a saying that I like to use when I start working with a singer who hasn't yet found his unique voice, and it goes, "embrace your flaws, because they are what make you special." (Or sometimes I'll substitute, "unique", "recognizable", "different", or some other similar word for "special".) The point here is I am not necessarily looking for the best singing, but rather the singing that best represents the character of the singer. If a singer has certain unique things they do, such as bending notes up or down at the end of certain types of phrases, doing little grace notes or turns here and there, turning monosyllabic words into multi-syllable diphthongs (any Reba fans out there?), etc., and I've got two equally emotional takes, one that is more vanilla and one that oozes the character of the individual, even if the elements that demonstrate that character might be things your classical voice teacher would scold you for doing, the quirky take wins every time, assuming, of course, it fits in the overall context of the composite performance.

I'm probably forgetting some other attributes, but, even with these, you can see there are a number of possible criteria. It probably goes without saying that all these criteria can be at odds with one another. One take may have some of these things going for them and others against them, another reverses things, still another seems to have everything going for it standalone but sticks out like a sore thumb when it is played in the context of what comes earlier and later, etc. Did I mention I consider comping to be tantamount to torture?

I'd normally get to a "where do we start?" consideration next. However, that is a complex subject, so let's defer on that for the moment and dive into the technical techniques that make that side of comping a relative breeze in SONAR 4. For you users of earlier versions of SONAR, in addition to the tortuous decisions you'll need to make, you'll also have to contend with a lot more technical efforts to audition takes spread across multiple tracks, including parallel muting and unmuting (or soloing and unsoloing) tracks, to accomplish the same sorts of things SONAR 4 allows you to accomplish with a single track with multiple layers and the Mute tool. If you do, or are considering doing, a lot of comping, I hope reading this next bit will give you some insight into just how much time, effort, and frustration can be saved after the upgrade to SONAR 4. In my experience, we are talking a major difference!

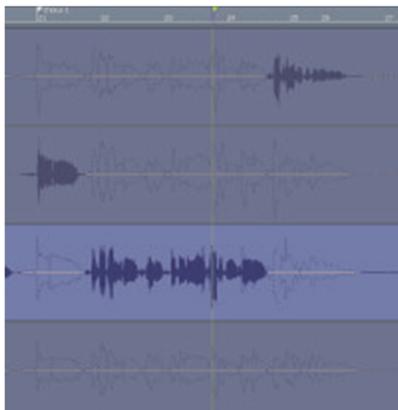


Prior to SONAR 4, about the only way comping could be reasonably managed was by using multiple tracks with a single take per track. If you wanted to audition a track with effects, as described in the previous section of this article, you'd have to have the same effects on each of those tracks or, in SONAR 3, route the individual tracks to a submix bus and put the track effects on the bus instead of the individual tracks. With SONAR 4, Cakewalk introduced multiple lanes within a track and a "Show Layers" mode to display any overlapping clips in separate lanes. Visually, it is similar to having multiple tracks within a track, but a key difference is that submixing of these lanes occurs before going to the track's FX bin. Furthermore, SONAR 4 has a new Mute tool, which includes the ability to "isolate" a clip, or region of a clip, within a track. This "isolate" capability is effectively a mutually-exclusive solo, but only within the track, where the clip you choose to "isolate" is heard, but any other clips overlapping it will not be heard. You're probably reading my mind by now, and, if so, you've guessed that this capability is extremely powerful when it comes to auditioning phrase-length clips for comping purposes. Here are the basic steps I use for comping a section of a song:

1. Maximize the track you will be comping to provide as much vertical space as you can for the lanes of the track.

2. If you haven't already done it, turn the "Show Clip Layers" mode on for the track so all takes are visible. (If you've been following the article thus far, you would have already done this at the same time as cutting the takes up into phrase-length clips.)
3. Select a time range of the song that corresponds to the section you want to comp plus a little lead-in and lead-out (I recommend setting the time range to measure boundaries), then use SONAR's "Set Loop Points to Selection" to tell SONAR to continuously loop on that range.
4. If you haven't already set the Click+Drag Behavior of SONAR's Mute tool to work in "Mute Entire Clips" mode, set that mode now.
5. Enable SONAR 4's Mute tool and Ctrl-Click+Drag over one lane of the track you are comping to isolate one of the takes as the starting point for your auditioning of takes. (When the Mute tool is operational, Ctrl-Click on any individual clip will isolate that clip, and adding the drag motion will isolate a range of clips.) 
6. Hit Play, and start listening to the current take.
7. As you hear things you do and don't like in the various phrases in the section you're comping, Ctrl-Click or Ctrl-Click+Drag to change the clip or clips you are listening to the next time the loop plays.
8. As you go through additional passes of the loop, start determining which phrases work better and which don't work as well, and use the isolate capability to try and hone in on the sequence of phrases you'd like to hear across multiple takes. As you start to refine your decisions somewhat, it may be helpful to delete clips you are sure you won't be using to narrow the visual field of considerations. (NOTE: If you don't like the idea of deleting anything, you can also move the clip to another track that is muted. However, another possibility, and the one I use, is to simply save an alternate cut of the project under another project name just prior to starting the comping, so I can always use that as an archive of deleted clips in case of emergency. In my estimation, comping is tortuous enough without having to look at clips you've already eliminated from consideration.) Keep this process up until you've either made all your final selections for that section of the song or come down to some decisions that either cannot be easily resolved at the phrase level (i.e. you might need to get into some finer splitting of clips), or which may require some additional experimentation to resolve (e.g. riding clip gain levels, trying some manual tuning of the clip, etc.).

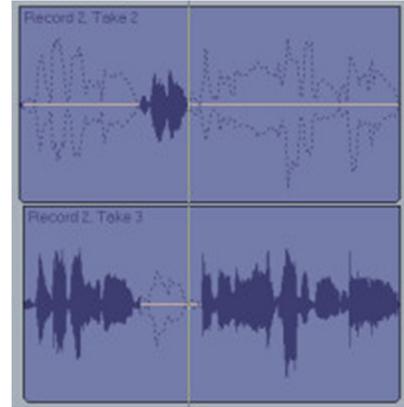
Of course, you don't have to keep looping through all of this. Sometimes it is easier to focus on a clip at a time. In this case simply taking advantage of SONAR's auto-rewind feature to start at the same place each time then play back a certain amount, manually stopping and restarting the transport, will make it easier to focus on individual decisions. However, keep in mind that the flow of the song, especially within a section, is also important. This is one place where not deleting clips you think you'll probably eliminate too early can be important since the best take in a one-on-one comparison against the other takes for that phrase may not be the best take in the context of the song's dynamic flow.



For those of you who are using the "one best take" strategy for comping, you will likely use SONAR 4's Mute tool with its Click+Drag Behavior in "Mute Time Ranges" mode. You can still use Ctrl-Click+Drag to isolate in this mode, but the isolation happens only in the time range over which you drag, rather than at the clip level. This can make getting the range of whatever it is you want to isolate, be it a phrase, word, song section, or whatever, a lot trickier, especially if the transport is rolling at the time and the screen is scrolling. However, if there aren't many sections to use from another take, this can save some time as there is no cutting takes down into phrase-length clips to do before starting the comp. You can always cut out the unneeded sections later, or simply bounce the lanes together using what is essentially clip automation to keep track of what gets heard at any point in time. For that matter, you don't even need to bounce things down at the track level as this

clip muting is honored at mixdown time, too. Just be careful not to accidentally change mute ranges once you've got what you want. While you can use Undo to go back to previous states, if you've made complex changes, and made new changes you do actually want to keep interspersed with those changes, it's not like you can just mute and unmute individual phrases like you can with the clip-level muting scenario we used above. You'll actually have to recreate the specific time ranges that represent those phrases.

It is worth mentioning that you can use clip-level muting in conjunction with time range muting. While I tend to prefer just cutting things up to allow visually getting rid of what I will be throwing out, this could be a way to slice things finer within a phrase without actually cutting up phrases. For example, let's say you're using the "no best take" comping strategy, and have things cut up into phrases, but you've reached a phrase where there are two candidates for the phrase in question, neither of which is perfect. The start of one is good, but there are some real nasty, unfixable pitch problems toward the end of the phrase. The end of the other is fine, but the start is emotionally anemic. While you could find a cut point within the phrase to slice things at the sub-phrase level and use clip-level muting, you could also keep both clips unmuted, and use time range muting within those clips to mute and unmute the portions of each clip you want to use. The example at the right of this paragraph shows how this might look in the case of a problem in the middle of the phrase that is better overall. That part of take #2 is used to replace the problematic section in take #3.



Okay, now we've talked about the technical techniques, but we've avoided talking about where to get started. I guess we should stop procrastinating.

It might seem that the obvious place to start is at the beginning. That is, begin comping a phrase at a time starting from the beginning of the song and moving backwards (or, if using the time range muting, reviewing the song from the beginning and moving toward the end.) That strategy can certainly work. It has the advantage of allowing you to easily make sure the all-important start of the song, which could make or break whether your listener keeps listening or switches channels on the radio dial or hits the Next button on a CD changer or MP3 player, starts out solidly. However, what if the takes you pick earlier in the song don't allow for building upon to make the dynamic level of the song rise as you approach its climax because the takes which come later in the song don't go up from there?

Another possible strategy is to start at the end. You guessed it, that means starting at the end of the song (or perhaps the point of climax if that comes before the end of the song), then working your way forward. This tends to ensure you can build up to what you've already chosen as there is an inclination to choose more powerful takes at the start, when you don't have a lot of context to set up the dynamic flow and you are just comparing individual takes. However, it gives you no better insight into what raw materials you will find at the start of the song than the starting at the beginning approach gives you into what will be available at the end of the song.

The net is, there is no right or wrong answer here. You may even wish to combine approaches, for example comping a section at a time, be it starting from the beginning or the end, but comping, within that section, from the opposite end of the section from the end of the song you started at. For example, if you are comping from the beginning of the song, you might start comping at the end of the first verse. This would tend to help you find something to build up to first within that verse, and work back from there, and any break between sections might give you a little leeway in matching energy levels to hit the start of the chorus. I've gone in both directions, and even started from the middle of the song (e.g. if the song climaxes at the end of a bridge, in which case the end of the bridge and the start of the section after that become pretty critical). I've sometimes even done a little bit at each end to start and worked toward the middle. The best I can suggest, beyond arbitrarily picking an end and going for it, is to somewhat familiarize yourself with the raw material before you start comping. Note any parts that may have particularly powerful material or particularly weak material overall, where you may want to either make sure something can be used or work around trouble spots, respectively. Then you can take those special considerations into account, and that may suggest to you where to start. For example, if you know that the takes at the start of the song are pretty weak overall, you may want to start comping there so you can at least get something workable. Your limited choices there may give some inkling as to where you can build to from there while maintaining a reasonable dynamic flow

of the song. Or, if there is a killer part that will absolutely make the song in another section, you may want to guarantee you get that. You can then move outward from there to make sure you get to and from that part in a way that works to highlight the part as it deserves.

The lead vocal can make or break a song, so, however fun or tortuous this part of the production process may be, it is critical to your end results. At the end of this part of the process, you will hopefully have a vocal that is compelling, emotional, and believable, while also making it clear who the singer is and having only easily resolvable technical problems.



Speaking of technical problems, while I prefer to separate the making of creative decisions from the resolution of technical problems, it is rare to be able to cleanly separate out these aspects of the project. It will often be the case that it is necessary to investigate a questionable tuning problem, try tweaking some consonant levels within a phrase, split some phrases and tweak timing to see how they can be merged together, and so on to make sure it is feasible to resolve any technical issues before making final comping decisions on some of the phrases in a song. In such cases, you may end up needing to use techniques we'll be addressing below during this phase. For that matter, you may even prefer to deal with everything at once, getting each section of the song solid as you go through it, rather than making several passes to address different problems. Again it comes down to personal preference. There is no right or wrong approach.

For those of you who have been following our audio examples, here is what I came up with after comping the section we've been using as an example. It does have one bit of microsurgery in it to repair a dropped consonant, but is otherwise a straight phrase-level comp:

["Baby" - Comped vocal with processing and automatic tuning](#)

Tuning

So here we are with a composite vocal that optimizes the tradeoffs between emotion, consistency, technical considerations, and lots more. However, if you've been following the process I've outlined to this point we have only glossed over pitch control issues, trying to make sure things were in the ballpark, and any issues would be correctable, but without having taken any real action to make such corrections.

Let's take a step backward, though. We've been using Antares AutoTune in automatic mode, and, if the pitch control within the comped vocal really is in the ballpark, it is quite possible that simply using the results of the automatic mode tuning will be good enough in some cases. This can be especially true when time is of the essence and the average music listener wouldn't notice the kinds of minor compromises in quality that would be made by taking this approach. In fact, I use this mode of tuning quite frequently for relatively quick demos, where I just want a few trusted colleagues to be able to provide feedback on a song or arrangement-in-progress, or am only planning to use the demo as a song demo, not a master recording.

If the results you're getting on tuning via that method (or without pitch correction software at all) suit you, then you may not even care to check out the rest of this section. However, maybe what is coming out of the automatic tuning isn't quite right, perhaps sounding like obvious tuning is going on or tuning a few things here and there the wrong way, or messing up the singer's vibrato. If so, this is where we will be getting into the finer controls available in AutoTune 4. Those controls make sure the pitch correction is doing what you want it to do, based on your own knowledge of the melody and what should be happening vocally, as opposed to what the software thinks is likely to be happening. Ironically, while this means putting more work into the tuning of vocals, the goal is typically to create results that sound more natural, and an experienced AutoTune user can indeed do that most of the time. This is because the software really has no concept of the melody, the singer's particular pitch-oriented inflections (e.g.

scoops, falloffs, etc.), the singer's specific style of vibrato, other than what it detects, which may be correct or erroneous. It takes knowing the melody of the song, along with the singer's style, to know what is truly intended.

Using the graphic mode of AutoTune allows us to see what is actually happening against concrete pitch references. It also allows us to see what AutoTune would do by default (i.e. if it were in automatic mode). However, it then allows us to make detailed tweaks to the actions AutoTune will take, or even ignore what AutoTune would have done on its own altogether and just tell it what we want it to do. Let's take a look at how this works. Before we do, though, you need to either delete AutoTune from the FX bin for the track you're tuning or bypass it. Otherwise any auditioning you'd do of the tuning would get doubly tuned -- i.e. once via your graphical mode changes and then with the automatic tuning changes set in the FX bin instance of the plug-in.

To use AutoTune in graphic mode in SONAR 4 (and any earlier versions of SONAR, or Cakewalk Pro Audio for that matter), you'll need to use SONAR's destructive audio effects from the Process/Audio Effects menu. The easiest way to get to this is to right click on the audio clip you want to process, and pick Audio Effects/Antares AutoTune 4 from the pop-up menu. If you're nervous about applying effects destructively, besides creating a backup of the project's CWP file (or WRK file from early versions of SONAR and Pro Audio) prior to tuning, you can clone the track with the comped vocal on it, archive it, and hide it. That way, if you change your mind about any of the clips you've processed later on, you can always go back to the cloned track, find the unprocessed version of the clip in question, and replace the processed version on your active track with that so you can start over on the processing.

One side note here is the ability to audition clips you will be tuning long enough to get through the entire clip depends upon your having set the SONAR Global Option on how long to audition commands to a length longer than the clip in question. The highest length of time you can set in SONAR 4 is 25 seconds, so I recommend setting that and leaving it that way. This means, though, that you can't graphically tune a clip that is longer than 25 seconds. This is not a big limitation, because it is much easier to tune smaller clips anyway when dealing with drawing or modifying the line graph AutoTune uses in this mode.

Once you have AutoTune 4 called up on a clip you want to tune, the first thing you'll do is set any general tuning criteria you may want to use for pitch tracking purposes, reference information, and default tuning criteria in the main AutoTune window. For example, in my test project, I set the Input Type to Alto/Tenor Voice, the Key to B-flat, and the Scale to Major. I also turned on Improved Targeting, which is a new AutoTune 4 capability that effects how AutoTune guesses about pitch changes in the presence of vibrato and certain other pitch-oriented inflections. The first three parameters I tweaked are self-explanatory. The last one has to do with AutoTune's behavior in areas where you are both changing notes and going into a note with vibrato. You'll have to experiment to determine whether you get better results with it on or off with your particular source material. I've found it to be helpful more often than not with the vocals I've processed -- it seems to give somewhat better accuracy in the corner areas between a slide up to a note and the onset of that note and the start of vibrato. While you are in AutoTune's main screen you can also set a variety of parameters relating to AutoTune's timing in tracking or adjusting pitch. You can also control whether or not to detune the pitch reference (e.g. to tune a vocal to an out of tune piano or to intentionally detune a background vocal to fatten the sound up when putting it against another background vocal that is either in tune or detuned in the other direction), how to deal with automating vibrato, and more. For most of my uses, I just leave these at default settings, because the defaults seem to work pretty well. I'll be making manual tweaks to the pitch curve anyway, so can do any customizations that way.

One other area I do sometimes tweak is the center section of the screen, where you can tell AutoTune to either Bypass or Remove notes from the scale you've selected. If you bypass the note, you are telling AutoTune to ignore it for tuning purposes (i.e. leave any note that would register at that pitch alone). If you remove it, you are telling AutoTune that, despite the note's being in the scale you've selected, it should not consider it to be among the reference notes to which it should tune pitches. The reason to do this would be because the note doesn't appear in the melody in the clip you're tuning and, thus, any note that might otherwise get tuned to that note was just too far off from some other note and being tuned in the wrong direction. Removing the note from the reference scale gives AutoTune better odds of tuning the note to the correct pitch. This capability can sometimes help with really shaky pitchiness in the original clip. This can be especially true if there are only a few notes in the clip and they are far apart enough that more extreme pitchiness get shifted to the nearest notes in the melody rather than the nearest, incorrect notes in the scale.



Once you've set any parameters in the main AutoTune screen, pressing the button to go to the Graphical Correction Mode gets you to another screen with a big empty graph taking up much of it. Most often, the next thing you'll want to do is press the Track button, indicating you want AutoTune to track the pitch of the clip, then hit Audition in the SONAR plug-in dialog for AutoTune, and let the clip play through at least once so AutoTune can track it. After stopping the auditioning, you should see a line graph appear in the graph area, showing the pitch detected by AutoTune for the original (i.e. untuned) clip. At left, you will see the pitch curve AutoTune has detected for the following short, as-yet-untuned vocal clip from our comped vocal:

["Baby" - Untuned word "baby" from comped first verse of composite vocal.](#)

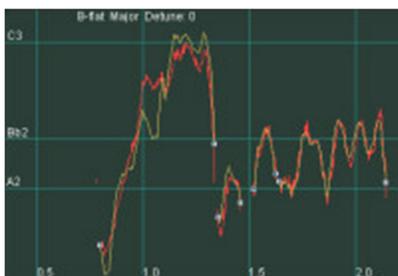
This pitch curve shows you what the vocal is actually doing (i.e. short of anything AutoTune can't resolve on the pitch front -- this can happen with voices with a high degree of noise and/or harmonic content, such as a very growly voice). However, it doesn't give you any clue to what it should be doing. To get that, you'd need to know the melody of the song (or the harmony part in question if this is a background vocal). For reference, here is the melody of the measure of the song that contains the short vocal phrase we are tuning here:



We are only tuning the first word in the above measure. If you compare this to the pitch curve AutoTune detected, you will note some decided discrepancies. For example, the first part of the first syllable in the word "baby" slides up from roughly A-flat to try and get to the C at the start of the melody, but doesn't quite make it to the C, starting out flat instead. Then the slur down to a B-flat in the melody behaves more like a slur down to an A-natural in the performance. Finally, the B-flat in the melody for the second syllable of the word slides up from A-natural to B-flat, then goes back down to A and up to B-flat twice, then finally goes back up to B-flat, wavering around it a bit until the end of the note, at which point it falls off to A.

This may sound like a pretty horrible performance, and, in fact, it is pretty darn shaky. (Why do you think I picked this example? It's no fun tuning something that is already perfectly in tune.) However, the melody alone doesn't tell us enough about what the singer intended here. If we tried to tune too precisely to the written melody, we'd get something that sounds like this:

["Baby" - Word "baby" tuned precisely to written melody.](#)



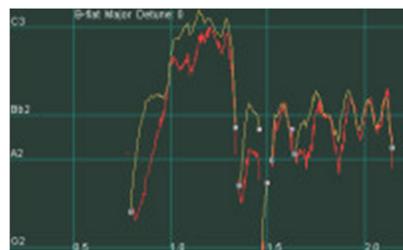
Pretty awful, huh? Yes, it sounds extremely robotic, and this is one key thing that makes synthesizing realistic human vocals such a difficult endeavor. Real singers don't sing notes perfectly according to a written melody. They slide up to notes, slide between notes, do falloffs at end of notes, use vibrato, do little turns here and there, and so on. AutoTune attempts to take this into account. To check out what AutoTune would do with this particular phrase without our help, we can simply press the Make Auto button, and AutoTune will draw a pitch correction curve using the parameters we've set in the main AutoTune page (i.e. where automatic mode parameters were set such as scale). In the screenshot at left, you can see that AutoTune decides the singer was probably trying to slide up from a G (i.e. since A-flat isn't in the B-flat scale, and the starting point of the slide was closer to G than A), then hit a B-flat briefly enroute to the C of the first note in the melody. It then decides that second note (the second half of the

first syllable of "baby", which we know should have been a B-flat) should have been an A (and, in fact, the singing was sufficiently off tune to hit an A reasonably solidly), then the last syllable should move back and forth a bit from a to B-flat before ending in some vibrato on the B-flat and the final falloff back to A. Here is what that automatic tuning would sound like:

["Baby" - Word "baby" automatically tuned to a B-flat scale.](#)

This might be a good time to re-mention that the various parameters on the main page of AutoTune can have an effect here. In fact, I mentioned that the "Improved Targeting" option sometimes makes things better and sometimes makes things worse. This is one of those cases where it makes them worse. If I turn off the Improved Targeting here, AutoTune slides directly up to the C on the first note instead of making a stop at B-flat. The Improved Targeting likely noted that the flat part of my first note was sufficiently flat, and held out long enough, that I probably intended there to be another note, such as a grace note. Thus, it went against its default instincts to tune it to the closest scale pitch and added that extra note instead. In fact, it might normally have been right, as my stylistic instincts as a singer probably would have me hitting a B-flat then sliding up to the C, almost like a grace note then slide. However, I was sufficiently sloppy in this case, that, if that was indeed what I intended (and it was probably early enough in the day on a day when I hadn't warmed up that I was probably just winging it without much of any intentions), I didn't come very close.

Let's take a look, though, at what tweaking AutoTune's automatic mode parameters can do, if used with a bit of knowledge on intent, though. We know that there are only two notes in the melody, the C and the B-flat. I also indicated my inclination as a vocalist would have been to slide up from the B-flat to the first note, rather than the roughly A-flat my sloppy singing yielded in this case. I can also tell you I would have used a little vibrato, and would have done some sliding between notes. However, the sliding here would have been intended to be mainly on the beginning and ends of notes, and between notes, not inside a note, such as what appears to be there on the last note toward the start. Also, any grace notes I might have put before the first note would have been short, not as long as is evidenced by the flat portion of that note just before it gets to the main pitch. With this in mind, I'll go back to AutoTune's main screen, have it remove all notes from the target scale except the B-flat and C, and turn off Improved Targeting.



Back in graphic mode, hitting the Make Auto button again yields the tuning curve at left. Just looking at this, it looks much more like what I would typically have in mind when singing this line. While there is still a slide up from my original A-flat, it goes quickly to a (roughly) B-flat grace note in the part of the word where I would expect to have done it, then goes up to the C by the part where I was just singing extremely flat. The second part of the melisma on the first syllable is now hitting the note intended by the melody (and what I would have sung had I been doing better). The variations on the final note are still a bit wide near the start of the word, but move to reasonable vibrato levels after that. Let's hear how it sounds:

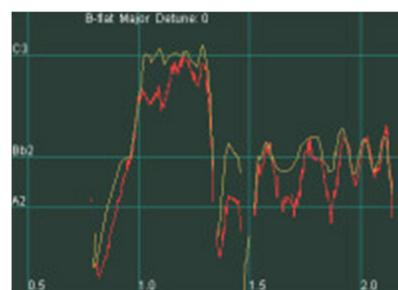
["Baby" - Word "Baby" automatically tuned to only the notes in its melody.](#)

Believe it or not, this is pretty darn close to what I intended, despite how different it is from what I actually sang in whatever state of voice I was in that morning. If I were pressured for time in getting a demo together, I'd probably call it close enough and move on. (I definitely would not have said the same for any of the earlier attempts because this one was just sung too badly to tune well with the more general tuning parameters set.) However, let's assume time isn't as big an issue as end quality for this production so that we can look further at AutoTune's graphic tuning capabilities, which I've barely used to this point other than for seeing what I'm hearing.

When I listen to the sample with this particular set of automatic tuning parameters applied, I hear a few artifacts in the tuned result. These are the sort of thing the average listener might not notice -- in fact, similar, and even worse, artifacts make it into recordings on the radio fairly frequently, and some recordings, especially in the pop/dance genre, have even featured much more severe artifacts as part of their sound. This isn't a dance recording, though, and I want as natural-sounding a vocal as I can possibly get given the source material. I'd even like to fool golden ears types who can spot "AutoTune artifacts" at a hundred paces, though I suspect that is an unachievable goal. Thus, I'll settle for making the result sound as natural as I possibly can to my ears. (For the record, I've been using AutoTune for quite a long time now -- since V1.3 -- so I've got a fairly good idea of what it does when applied transparently and not so transparently.) I'll assume if I can do that, it'll fool almost all "normal" listeners (i.e. anyone who isn't fairly intimately familiar with the artifacts pitch correcting vocals can create).

Where am I hearing artifacts? Even if you can't hear the artifacts yourself, the locations will probably come as no surprise, if you look closely at the tuning curve that was applied here. The main places are on the slide up to the first note and toward the beginning of the last note, where you see the extended semi-tone dip in the original pitch curve. It wouldn't surprise me, from looking at the graph, if there were also one on the second note of the melisma, where AutoTune had to correct my pitch by an entire half step. However, I'm not hearing anything even mildly offensive there, so either AutoTune did a very transparent job of this fairly severe shift or my ears just aren't quite savvy enough to notice whatever artifacts may be there.

My experience with AutoTune tells me that the likely reasons for the artifacts I'm hearing in the first case are because AutoTune is trying to create a short held note where there is none. There is only a slide. In the second case, AutoTune is trying to remove a very short held non-scale note where there effectively is one, but only due to my pitch gyrations of that morning. These somewhat extreme manipulations go against the grain of how the original pitch curve ran (and are based on my instructions to AutoTune as to what to accept -- it wouldn't have tried something this extreme with its default settings). They force AutoTune to try to do too much too quickly, thus generating the artifacts. I will try to eliminate the artifacts by using my knowledge of the song, and my intentions as a singer, to try and make as good of tradeoffs as I can between my intentions (which AutoTune actually hit upon pretty accurately) and the reality that I didn't achieve my intentions in the performance. I will do this while also paying close visual attention to the original pitch curve to try to make relatively smooth moves while still taking my intentions into account. In particular, I will smooth out the slide up to the first note, leaving only a trace of the grace note on B-flat. The goal here is to get rid of the artifacts that came from the forced, longer grace note the previous attempt had added. I will also recognize that my pitch near the beginning of the last note dipped. Instead of trying to pretend it didn't, or assuming it was actually meant to go an entire half step down, I will draw a change in pitch, but minimize the amplitude of that change. I will also keep in mind what comes later on in the vibrato to try and make it feel more like just an early slow vibrato that speeds up over time. I did these curve modifications using AutoTune's line tool, as I find dealing with straight lines easier to do with a mouse than freehand drawing. However, the changes could also have been made with the curve tool. Have a listen to the result:



["Baby" - Word "baby" with manual adjustments to smooth out previous controlled automatic tuning.](#)

To my ears, that is just what the doctor ordered. I'm not sure I could tell the difference between this result and if I'd just sung it right in the first place. Whether I could or not, though, if I listen back to this and compare it to the original untuned clip, or the default automatic one I would have ended up using if I'd just use AutoTune in automatic mode, there is a major improvement.

Though the steps we took in this section may seem like an extraordinary amount of work to fix what turned out to be a single word, most of the steps taken above would either not have been taken in real life or would have happened extremely quickly. A little experience with graphic tuning, plus a decent ear for pitch, goes a long way toward being able to predict issues from simply looking at the tuning curves. This often allows adjusting the curves before even auditioning the result, though it is also pretty common to audition an initial result then refine the curves with a few additional passes. This approach can be especially useful in longer phrases and in cases when there are tradeoffs

between trying to come close to what the singer intended (or what you think she intended if you're not the singer) and the likelihood of artifacts if you try to enforce things too closely. In this particular case, I can't imagine tuning this word would have taken more than a minute, if even that much. I probably would have removed notes from the scale right from the start, and would have just drawn any other changes from whatever Make Auto generated at that point.

To get a feel for how far we've come to this point, here is the full first verse, comped and tuned:

["Baby" - Comped vocal with manual tuning.](#)

Total tuning time for this verse was about 5 minutes. Here is the same vocal against the instrumental tracks:

["Baby" - Comped/tuned vocal with tracks.](#)

Okay, so we've cheated a little. The original temp tracks have had their instrument sounds replaced with new, higher quality sounds. But you get the idea.

Basic Editing

Once I have a vocal comped and tuned, I like to make what might be considered a quality control pass over the composite vocal, cleaning things up and looking for any issues that may need to be fixed or smoothed over. There are various types of issues that could arise, from a dog's bark or cough between two phrases, to too abrupt or otherwise awkward transitions between phrases, to minor diction issues, to rhythm or other phrasing problems.

Most truly major issues would have been addressed in the comping process. For example, a dog's bark in the middle of a sung word would likely make that particular take of the phrase, or at least the word, unusable since it would be difficult, if not impossible, to remove without affecting the vocal itself. However, we are generally going for passionate performances, and our experience and judgment will tell us that certain more minor issues can be overlooked at comping time, if it means getting a more emotion in the performance, because we can fix, or at least hide, them later. Of course, we can also fix them while comping. However, fixing certain types of problems may require shifting from right brain work (i.e. evaluating performances based on emotion and other aesthetic issues) to left brain work (e.g. figuring out how best to smooth a transition between the start of a word and an ending consonant that has to be grafted on from another take because it was dropped in the take that was otherwise best). Thus, unless there is some question as to the feasibility of fixing a particular problem, I tend to lean toward keeping comping as far toward the aesthetics as possible, and saving the technical work for a second pass. It's like shifting roles. During the comping you are mainly a producer, even if you might need help from the engineer from time to time. In this phase, though, you are mainly an engineer, even if the producer's judgement may be needed occasionally.

We'll discuss a few more complex techniques a bit later, but most of what we'll be doing in this phase of the production process simply recognizes that we are trying to make a performance composed of pieces from multiple takes appear to be one seamless performance. To use an analogy, even a single wall in a home will often be put together from multiple pieces of drywall. To make it look like a single wall, tape, plaster, and sometimes "mud" (I'm not sure of the actual name for this, but I'm talking about the stuff that they spray over a wall to make it look textured) are used to hide the seams between pieces of drywall, then everything is painted to further smooth things over and polish the results. What we are trying to accomplish here is very similar in intent -- i.e. to try and make sure others don't notice our "seams".

Some of the minor issues we'll look at here include:

- Keeping background noise out of the way and avoiding obvious changes in background noise level

- Smoothing transitions when phrases have minor (i.e. non-problematic) overlaps
- Managing breath noises

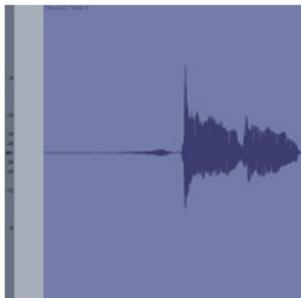
The tools we will use include SONAR's slip-editing facilities, real-time fades and crossfades, and SONAR 4's new overlap cropping tool.

Those of us who have non-soundproofed home studios have all experienced times when the neighbor's dog (or your dog) barks, a car alarm goes off, the baby cries, a toilet flushes, the phone rings, etc. Then there are the relatively constant, but hopefully reasonably low-level, home studio noises like the fan and disks from your DAW. Even those of you lucky enough to have a soundproofed studio or vocal booth will have to contend from time to time with coughs, lyric sheets shuffling, a singer's hitting the mic stand, and so on.

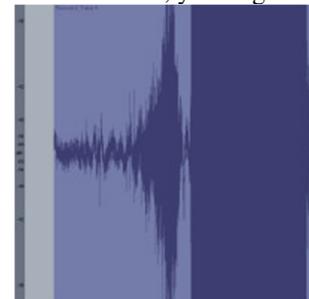
For our purposes, there are three types of background (i.e. non-desirable) noises. One is a constant level of background noise that just can't be helped due to the specific studio environment. Hopefully it is low enough to be masked by the vocal when the singer is singing, but it can still be a concern in between vocal phrases. The second is stray noises, of the type that are loud enough to be heard, and which occur between vocal phrases. The third is a variant of the second -- i.e. stray noises that occur during vocal phrases, and which are not masked by the singing. We will ignore the third type here because there generally isn't a lot we can do about this type of noise, and hopefully the comping process has eliminated noises in this category. We will address the other two types.

In the case of background noise that is just part of what you get in your studio environment, if that noise is of a constant type, it is possible that it may be able to be removed via noise reduction processes. This would be accomplished using third party plug-ins or other audio tools which take a "noise print" of the noise, then attempt to remove that from the audio. If the noise is intrusive enough to be noticed even when there is singing going on, this may be a necessary evil -- an "evil" because even the best noise reduction can still adversely affect the desirable part of the audio signal. Thus, there is often a tradeoff between removing sufficient noise and avoiding noticeable artifacts in the vocal itself. We won't be further considering that approach here since it is generally unnecessary when dealing with vocals, but it just might be the ticket in some situations.

The approach we'll take with constant background noise that is masked by the vocals, but too apparent in silent sections, is to simply edit it out in a manner similar to how a noise gate works, but with much more control. You may be wondering, if this approach is similar to a noise gate, why not just use a noise gate? Of course, that is a possibility, and may be a good route to go if time is short. The biggest problem with using a noise gate on vocals, though, is they can't consider the context of low-level signals (i.e. beyond any side chain possibilities), just the level of the signal. Thus, they are indiscriminate about what they cut out, treating desirable low-level signals the same as they treat undesirable low-level signals. Using SONAR's slip editing capabilities allows us to decide what we keep and what gets removed. Just as significantly, it allows us to decide how we will handle the transition between silent and non-silent areas.



Let's take a look at an example. The screen clip at left shows the short vocal phrase that begins the first verse of the song. You may recognize that there is some "silent" area, followed by a breath, then the actual vocal phrase, in this case, the word "baby". From looking at this vocal clip, which is scaled to SONAR's default level, you might think there is no significant noise prior to the breath mark. However, let's scale the waveform up so we can see lower levels of the dB scale. The screen clip at right shows that very same clip with the waveform scaled to make lower levels of the dB scale more apparent. The scaled waveform suggests there is background noise with an amplitude somewhere in the neighborhood of -54dB. This noise happens to represent what is commonly heard in the background in my recording environment -- basically computer fan noise. Have a listen:

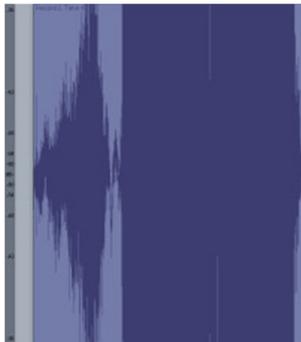


"Baby" - Raw noise at start of vocal phrase.

If you're listening on cheap computer speakers, or other low-end consumer speakers, you may not have even noticed it, especially if your own computer is making its own fan noises. Even if you're listening on decent studio monitors, while the noise is audible, it may not seem terribly offensive, and would almost certainly be masked by the instrumental tracks in mixes for most modern commercial genres. Thus, you might be wondering why we should concern ourselves with this relatively low-level noise. That is a decision each of us must make, as there are tradeoffs between the time taken to deal with the noise, and how apparent it will be in our end product. However, there is another consideration that comes into play here. In particular, what you heard in that last audio clip was the noise from the raw vocal track, with no processing on the vocal. Now let's have a listen to that same noise once the processing we've been using for auditioning the vocal is applied:

"Baby" - Noise at start of processed vocal phrase.

I'd venture you probably heard that processed noise even if you're using cheap computer speakers and your computer fan is running. If you were listening on studio monitors, though, words like "annoying" and "offensive" may well have come to mind.



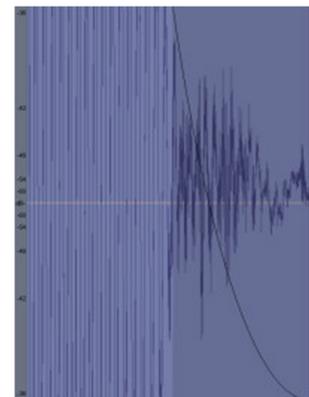
Now let's use SONAR's slip editing to trim the noise back to the start of the singer's breath. I'll leave some of the silence just before the breath in the mixed audio clip so you can hear where the noise would have been:

"Baby" - Noise before processed vocal phrase trimmed back to start of breath.

Note that I primarily used visual clues, after moving in zooming even further in on the waveform's scale, to determine where to move the starting point of the slip-edited clip. I was looking for a marked difference in pattern of the waveform, not necessarily a dramatic rise in level. After all, the start of the breath is more or less buried in the background noise. While I used visual cues to pick the editing points, I used SONAR's ability to play only what is selected (Shift+Spacebar) to check that my editing points sounded okay.

Because of the singer's breath at the start of this clip, the transition between pure silence and the start of the clip was not noticeably abrupt, at least to my ears. However, I also left a little bit of pure silence after the end of that previous audio clip, in a location where there would be a fairly tight transition between this phase and the next, which may be coming from a different take of the performance. If you listen closely on good speakers, you may have heard a slight artifact of the abrupt transition from the end of the clip to silence. Now, in the context of an actual mix, odds are this artifact would be masked by other tracks. However, for the sake of example, let's say this was going to be an a cappella mix. In that case, it might be desirable to effect a smoother transition between the end of the vocal phrase and the silent area prior to the next phrase.

We will use SONAR's clip fade-ins and fade-outs to deal with transitions between silent areas and clips, and between clips and silent areas, respectively. In this case, we will put a fade-out on the end of the clip. SONAR has three types of fade-ins and fade-outs: A linear fade fades smoothly, in a straight line, from the start to the end of the fade. A fast fade starts to change the volume quickly at first, then tapers off the rate of change toward the end of the fade. A slow fade starts to change the volume slowly at first, then increases the rate of change toward the end of the fade. In this particular case, auditioning all three types of fade-outs demonstrated that a fast fade was the only way to prevent the artifact at the end of the clip near the transition point. Looking at the waveform of the clip might have predicted that anyway, as there is a dip then rise in the waveform near the end of the clip. However, it is still a good idea to audition any fades to make sure the starting point of the fade, along with the fade curve being used, does not adversely affect the desirable part of the audio, in this case the end of the word "baby". Have a listen to the same vocal clip with an end fade into silence:



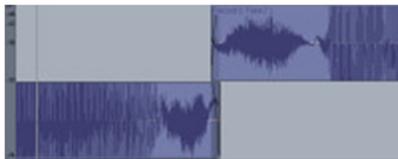
"Baby" - Trimmed/processed vocal phrase with fast fade-out.

As a rule of thumb, you can often get away with a more rapid fade-out on a word that ends in a vowel than on a word ending in a consonant. In the case of a word ending in a consonant, the rapid fade is more likely to harm the diction, lowering the consonant level too quickly. If a rapid fade-out is needed in that case, it may need to be started closer to the end of the clip.

While fade-ins and fade-outs may not be absolutely necessary in many cases, since instrumental tracks will be covering up low-level noise and artifacts anyway, I tend to use them on all clips as a matter of course. Call it superstition, or being anal retentive, or whatever you like, but somehow it just feels cleaner to me, and it is very quick to do when reviewing the composite vocal at this level since the right type of fade for each situation will tend to become intuitive after awhile. Also, if you set your default fade-in and fade-out curves to be the type you most commonly use, fading most clips will simply be a matter of dragging the appropriate corner of the clip to pick your fade ending or starting points. SONAR's default of fast fade-ins and slow fade-outs seems to work most of the time for me.

Everything I just described for dealing with constant background noise between vocal phrases also works for dealing with stray noises (e.g. coughs, burps, barks, meows, etc.) between vocal phrases.

So far we've discussed clips that start or end in silence -- i.e. they do not overlap the clips that contain the prior or subsequent vocal phrases. What about tighter transitions, where there is some overlap? For the moment, we'll talk only about minor overlaps, where there isn't a danger of mangling the phrase. For example, this might occur if the earlier phrase is significantly longer in the best take for that phrase than the same phrase in the take from which the later phrase comes, where one word would cut off another. In cases where these problems exist, if they weren't weeded out in the comping, we will have to use more complex editing techniques discussed later in this article.



For these simple cases of overlap, say where the we are only talking about background noise overlaps, breath and/or background noise overlaps, and other relatively easy-to-handle overlaps of less important content (as compared to the actual vocal phrase itself), SONAR provides two capabilities which can be helpful. The first is automatic crossfading, where, if you overlap one clip with another on the same track, the region in the overlapped portion of both clips will automatically be crossfaded between the two clips. That is, the earlier clip fades out in the same space in which the later clip fades in. You can set the default fade curves for the fade-in and fade-out directions in a crossfade to be any combination of linear, slow, or fast fade-ins and fade-outs. For example, SONAR's default is for a slow fade-out and fast fade-in.

The basic idea of a crossfade is to blend the transition between two clips as seamlessly as possible. The key thing the crossfade achieves is trying to even out the volume in the overlapped regions by backing off the level of one clip while incrementally introducing the other. In cases of crossfades during background noises, breath noises, and even some consonants (e.g. sibilant ones such as the letter "S"), the transition can be pretty seamless, to the point of being totally unnoticeable. We will talk later about using crossfades in more adventuresome ways, but these simple cases tend to be quick and painless.

New to SONAR 4, is an overlap cropping tool. This tool, which takes the form of a special cursor, appears when you place the cursor between two adjacent track lanes. It is used to align the ending point of the earlier clip to the starting point of the later clip. Moving the cursor with the left mouse button depressed moves the crop point for both clips. If automatic crossfading is not enabled, this cropping action butts the end of the earlier clip right up against the start of the later clip. If automatic crossfading is enabled, however, SONAR creates a slight overlap between the clips surrounding the cropping point you indicate with the overlap cropping tool, and this overlap gets an automatic crossfade between clips.

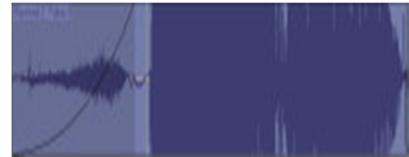
The final area we'll be discussing in this basic editing section is management of breath noises. Before I get into the details, though, I'd like to mention one important point for those who may be new to the notion of editing vocals and

who may be thinking this stuff we've been talking about is great and the obvious thing to do with breath noises is eliminate those pesky noises, treating them like any other undesirable noise. To quote a current hit country song, "don't ask me how I know," but this is decidedly not a good idea, especially for lead vocals. In particular, doing so pretty can pretty much single-handedly suck the life out of an exposed vocal. As unintuitive as it may sound at first, the types of breath noises we make can be a big part of our overall expressiveness.

While removing breath noises altogether may not be such a great idea, leaving them *au naturel* is often not the best course of action, either. For one thing, there are the semi-obvious cases where a singer runs short of breath and gasps instead of inhaling more normally. Comping may be likely to eliminate some of that sort of thing (i.e. because a line that requires a gasp at the end of it has good odds of exhibiting some symptoms of lack of proper breath control earlier in the line). You may still end up with some of that, though, or with slightly less extreme breaths that are nevertheless a bit much for the context. Also, characteristics of specific mics and mic techniques (e.g. proximity effect) can sometimes exaggerate breaths beyond what you would hear if you were just listening to the singer in the room. Then there are times where even a relatively natural-sounding breath may be a bit too much for the track. Not to mention that effects can sometimes exaggerate breath noises.

Additionally, because we used breaths to help us determine where to cut up the vocal into phrases for comping, differences in phrasing between takes can create issues in this area when two different takes are combined in the composite take. For example, in one take a singer may breathe at the end of a phrase, then have a short pause, then start the next phrase, while in another take the singer may put a pause at the end of the first phrase then breathe just before the second phrase. If the first phrase in the composite track will come from the first take and the second will come from the second take, you could end up with two breaths. Another scenario might be where the first phrase comes from a take that holds an ending note longer than the second take does, and the first take's version of that phrase will be combined with the second phrase from the second take, where the singer has breathed while the note from the other take was still being held. This could create a scenario where a breath is needed, but there either isn't the space for the entire breath, or where neither take's positioning of the breath will work in the composite take.

Often the issue with an exaggerated breath is more about how quickly it ramps up to a strong volume level than where its volume level ends up going. If a singer breathes at the beginning of a phrase, a fade-in can be an easy way to deal with this. Whether to use a linear, fast, or slow fade-in will just depend on how severe the issue you are trying to deal with is. Check out the previous audio clip for the phrase we've been working with, then compare it to the following clip, which has had a slow fade-in applied to the entire length of the breath:



["Baby" - Slow fade-in on breath noise.](#)

In the earlier clip, the breath noise was pretty severe, to the point where it had gone beyond expressive and into the territory of possibly annoying. With this slow fade-in, though, expressiveness is maintained without the sudden onset of the breath. A further advantage is that the slow fade-in also keeps the level low in the portion of the sound where the breath noise and background noise are at relatively similar levels.

If the breath is at the end of a phrase instead, a clip-level fade-in obviously can't help deal with issues in the early part of the breath. It might be possible to split the clip and do a quick fade out on the end of the first one and then do a fade-in on the second, or you could use clip envelopes to create a similar effect. However, there is always a danger of affecting the end of the word before the breath. Perhaps it is because a singer who tends to breathe at the end of a phrase, rather than the beginning, may have better breath control in the first place, but I have found that I don't seem to have as much problem with breath noise that occurs at the end of a phrase as with breath noise that occurs at the beginning of a phrase. When there is a need for reduction of the breath noise at the end of a phrase, it seems that a fade-out is often sufficient. A fast fade-out, or just starting a bit earlier than the breath noise with a linear or slow fade, can often be useful to reduce initial breath volume levels.

When it comes to double breaths or other multiple breath issues (e.g. 1.5 breaths coming from a breath on each end of a comp split point where one of the breaths partly overlaps the other), the easiest thing to do is typically to slip

edit out one of the breaths, favoring leaving in the one that sounds the most appropriate for the track. A fade-out or fade-in may be necessary to try and get as much of the extra breath out as possible without destroying a word ending or beginning, respectively. A slight variation on this theme is when you have a breath at the end of one phrase and one at the beginning of the next phrase, where the space in between is just right for one breath, but you can't easily get rid of either without hurting a word ending or beginning. In this case, finding an arbitrary split point somewhere in the middle of the two overlapping breaths and using SONAR's overlap cropping tool with automatic crossfading turned on generally works very well as breath noises tend to be relatively easy to crossfade with one another without noticeable artifacts.

A too tight spot for the breath noise you have is only slightly trickier. Most of the time you can simply do a fade-in or fade-out after slip editing the breath to be short enough to fit the allotted space. This can be a situation where the rate of the fade matters more than in many circumstances since you'll be starting somewhere after the original breath was already in progress or ending somewhere before the original breath would end. A little experimentation will tell you which type of fade sounds best. If just trimming the end of the breath and fading doesn't give a natural result, another possibility is splitting the middle of the breath, then sliding one end toward the other so you still have the original beginning and end of the breath noise. You can crossfade the overlapping parts in the middle together, possibly using the overlap cropping tool, for a shorter breath with roughly the same contour. Of course, there are always more advanced techniques like time stretching (see below), but generally the above-mentioned methods are quicker and just as effective.

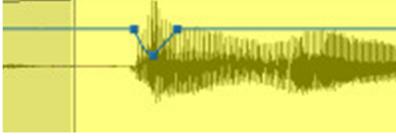
Assuming you've done a solid job in the comping and tuning processes, and have a composite track whose individual components flow well together, these easy basic editing techniques tend to be sufficient for dealing with a large percentage of the issues you are likely to encounter while trying to make your composite track feel like a single performance. However, in most comped vocals there will be a few problem areas needing special attention, and some more advanced techniques. We'll be covering some key situations and techniques in the next few sections of this document.

Editing: Fixing Dynamic Problems

Comping can be done at any level, but most of the time I prefer to avoid slicing things finer than at the phrase level to preserve the singer's phrasing, and continuity within each phrase. While some problems may require cutting things up down at the word level, or at an even finer level, I tend to consider that sort of thing a last resort. One issue that might be tempting to consider addressing by cutting things up finer, to grab individual words from different takes, is when there are micro-level dynamic problems, such as an over- or under-stressed word, a popped "P" or "B", a too hissy "S", etc. Sure, you could just grab the word that has the problem from another take that doesn't, but that could risk making the phrase feel a little less like a single performance. Thankfully, though, dynamic problems of this type can generally be resolved without resorting to lower-level comping.

In fact, we've already touched on this to a degree in addressing dealing with overzealous breath noises, where a fade-in or fade-out might be just the thing to resolve the issue. If the dynamic problem isn't at the beginning or end of a clip, and/or requires something more complex than a simple fade, clip envelopes can come to the rescue. For example, if you've got widespread sibilance problems in a vocal, you may want to employ a de-esser. However, if there are only a few stray instances, using a clip envelope to ride volume level on the clip to reduce the level during the offensive "S" sound won't take all that much time. Additionally, it may yield better results than the de-esser since it won't be affecting anything other than the one "S" sound (or however many of those you choose to address this way).

Let's look at one real-life example in our continuing work with the first word of "Baby". If you listen back to the previous audio clip, you may note a fairly exaggerated initial "B" sound. Perhaps I just popped my "B", or perhaps the "B" was exaggerated by the microphone's proximity effect. Whatever the cause, though, the *sforzando*-like effect it creates feels a bit harsh to me for this song. I might note that we've already been auditioning this track through a compressor, so, though compression or limiting certainly could be adjusted to help out here, the settings we've picked for generally applicability to the track don't seem to be doing enough in this case. Clip envelopes to the rescue!



If you look at the waveform visually (I've changed the color of the clip in this example to make the clip envelope show up better), you can see there is a noticeable peak near the initial "B" in the word. Inserting a clip envelope for gain level, I will first create a node right around where that peak exists, then two additional nodes before and after the peak where things more or less get

back to normal. I can then drag the node at the point of the peak down by approximately 6dB, which roughly halves the apparent volume level at that point. (This setting is probably a bit extreme, but I wanted to use a value that would be noticeable enough for our audio examples.) From there, it is a matter of auditioning the results and playing around with the curve types in the segments between envelope nodes to try and achieve the most desirable result. Here I ended up using a fast curve to get from the start of the word to the peak, so the initial popping would be rapidly controlled, then used a linear curve to return gradually to the point at which this dynamic tweak ends. You can hear the result here:

["Baby" - Clip envelope used to reduce "B" popping.](#)

Of course, this is all subjective, but this feels better to me given that I tend to think of this song as a bluegrass lullaby, and the popped "B" might well wake the baby up. Okay, so maybe that's a bit of an exaggeration, but the perhaps-too-subtle point is that the idea is not just to make these kinds of micro-level adjustments because we can, but to benefit whatever the desired result may be in the context of the overall song and recording.

Sometimes a dynamic change needed within a phrase requires more than just a volume adjustment. For example, sometimes the reason for the shift in volume was something that brought along with it a shift in timbre, and just raising or lowering the volume will still leave the result sounding somewhat out of place. For example, the timbre might be overly bright or overly dull sounding compared to the rest of the phrase. When this happens, it may be possible to couple the clip gain riding with some destructive audio processing to address the timbral side of the picture. For example, in the case of an overly dull word in an otherwise good phrase, it may be possible to process just that word with a harmonic exciter (e.g. Cakewalk's Project5 HF Exciter, the DSP-FX Aural Activator, PSP Audioware's MixTreble) to brighten it up, or an equalizer may be able to be used to do a similar thing. If the word is overly bright, then you can go the other way. While I'm not aware of many "harmonic borers" ("harmonic de-exciter"? "harmonic depressors"?), I have had good results the Smooth/Enhance plug-in from Sony's Sound Forge. Setting positive values works like a harmonic exciter, while setting negative values makes the sound less edgy. Of course, some EQ can be used to accomplish a similar end result, too. This kind of change works best if you can isolate a word or syllable and make a smooth breaking point. However, if that is not possible, it may be possible to snip the phrase into smaller clips around the portion of the phrase you want to tweak, slip edit the ends of the various clips to create overlaps on each end of the parts you'll be tweaking, process only the clip you'll be tweaking, then adjust the crossfade regions between clips to try and get a smooth, hopefully unnoticeable, transition.

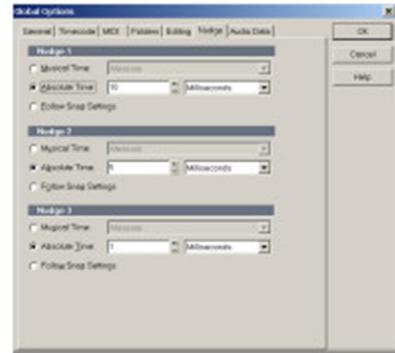
Dealing with dynamic tweaks at this level can be a bit tedious. However, the results can be very worthwhile if it means you can use less aggressive compressor/limiter settings since the compressor is only being used for overall smoothness, rather than fixing problem areas. This can let more of the dynamic of the original vocal come through, thus preserving more of its emotional punch.

Editing: Fixing Phrasing Problems

While comping is intended to achieve the most believable, most emotional performance, and phrasing plays a big part in that, sometimes there are too many tradeoffs to be made in the comping process, and phrasing problems will make it through to the composite vocal. For example, perhaps other takes had too severe of pitch problems to allow transparent enough correction with AutoTune, or the singer sang the wrong words in too many of the takes, and you're left with a take that is mostly okay, but just isn't as good as you know it could be in the phrasing department. That might be a good time to get the singer back in to overdub the problem phrases, or to sing them again if you happen to be the singer. However, sometimes that isn't practical, and you are faced with the choice of settling for results you know are non-optimal or seeing if you can work some studio magic to improve those results.

For our purposes, there are two types of phrasing problems: rushed/dragged phrases and phrases which are sometimes rush and sometimes drag. When an entire phrase is rushed or dragged, the easiest way to deal with that is to select the clip(s) containing the phrase, and move them forward or backward in time. This can be done by dragging with the mouse, which may be the best route if there are visual references you can use to align the vocal with, for example, the drums. However, if the timing differences are small, an even easier way might be to use SONAR 4's new Nudge command.

The Nudge facility in SONAR 4 allows easily making very small changes to the clip's starting position based on values you've predefined for the amount of time by which a clip should be "nudged" depending on which key you press. Pressing the "5" key on your keyboard's number pad brings up SONAR's Nudge Options dialog. (These can also be reached from SONAR's Global Options.) There are three degrees of nudging that can be set, and which one is used in an actual nudge operation depends on which key on the numeric keypad you select. Going down the number pad, the "7", "4", and "1" keys nudge whatever is selected left (i.e. backward in time), and the "9", "6", and "3" keys nudge the selection right (i.e. forward in time). The amount of time for each row of keys is set in the Nudge Options. Various musical and absolute time units can be set, and you can also force nudging to snap to your grid settings. For our purposes, though, since we are talking about nudging audio clips, I prefer setting a few different levels of absolute time, and generally set my options to 10, 5, and 1 milliseconds. That way, if I think I need a fairly large move, I might start by nudging by 10 milliseconds in the direction I need the clip to go. As things get closer, I might move down to 5 millisecond units, then, when I am almost there, I may go back and forth trying options with 1 millisecond nudges. Nudging is most useful in conjunction with auditioning the clip you are nudging in the context of the mix. Looped playback can be useful to avoid having to keep starting and stopping the transport.



Let's take a listen to what some simple phrase-level, in fact verse-level in this case, nudging can do to the feel of a vocal track against the instrumental tracks. This first audio example is the original vocal, with much of the tweaking we've discussed to this point applied, but prior to doing any nudging:

["Baby" - First verse before nudging.](#)

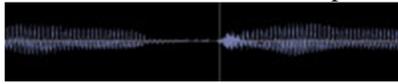
It sounds pretty reasonable, but there is a subtle feel of urgency in the vocal which may rub the wrong way in a lyric that reads more like a lullaby, bluegrass or otherwise. That feel seems to pervade the track, perhaps because the vocals were tracked against pre-recorded tracks, rather than with all the visual cues that go along with a true live performance with multiple musicians playing together, or perhaps just because the singer was in a rushing mood that particular morning. Whatever the reason, though, it might be useful to experiment with nudging the vocal backward a bit, to see if that improves the feel. This next audio example was created by selecting the entire first verse vocal, then hitting nudge twice with that particular nudge setting set to 10 milliseconds, thus delaying the vocal by a total of 20 ms against the instrumental tracks:

["Baby" - First verse vocal nudged 20ms later.](#)

To my ears this is an overall improvement. However, I'm not ready to commit to it yet as my sense is some phrases would probably benefit from more or less delay than other sections. Perhaps some phrases might even need slightly different amounts of sliding within themselves. If you don't mind, I'll procrastinate on this decision until I'm finished writing this article and have more time to concentrate on the vocal editing.

That said, this does bring up a further consideration: While nudging an entire phrase is simple and convenient, sometimes the timing of the phrase is just plain off, where some of it may be perfect, other parts may be rushed, and still other parts may be dragged. When this happens, about all you can do is cut up the phrase into smaller clips, then slide or nudge the individual components as needed, and use the various techniques we discussed above in the Basic Editing section to smooth over any newly created transition points. Being able to do this easily depends upon finding convenient points at which to slice up the phrase. Sometimes there will be slight spaces between words that provide

convenient slice points, and sometimes one word will flow smoothly into another. Sometimes there may be no between-word locations, but a particular word might have a good point just between two consonants or between a vowel and certain types of consonants. The clip at left shows a point in the transition between the end of the word "every" and the beginning of the word "deed", where there is a clear slice point just before the first "D" of "deed".



We'll talk more about dealing with cases without clear slice points when we get to the "Microsurgery" section below. If there are good slice points, though, it is pretty easy to just slice things up, nudge each individual clip around as needed, smooth over transitions as we've done previously, and go on our merry way.

Of course, this is assuming that everything within each smaller clip is off in the same direction and by the same amount of time. If not, there are two possibilities for dealing with it: "microsurgery" (see below) and compromise. I would tend to lean toward compromise unless the results you get by just making do as best you can with conveniently sliced clips just doesn't cut it in this case. The reason is that it can be very tricky to slice things up and manipulate things when there aren't convenient slice points. You can spend a lot of time trying to deal with the kinds of artifacts that can easily arise when, for example, you try and connect two words that should flow together, but need to be moved in opposite directions due to bad timing in the original phrase. We'll get around to dealing with such cases momentarily.

In the interim, though, it is worth mentioning that having good slice points doesn't guarantee you won't have to deal with some of the same kinds of issues we'll be dealing with in the Microsurgery section. For example, let's say you have two words that come together, but with an excellent slice point between them. You have to cut between the words, then move the first word earlier in time and the second word later in time. That's easy enough to do. However, now you may have too much space between words, creating an awkward pause that may even be worse than the problem you were trying to address. If there is any saving grace, it is that, though you may still be forced to resort to compromise or some of the more advanced techniques we'll be talking about in the Microsurgery section, having good slice points does at least improve your odds for avoiding artifacts when doing things like time stretching (e.g. to fill in the gap between these two words) then splicing things back together.

Editing: Microsurgery

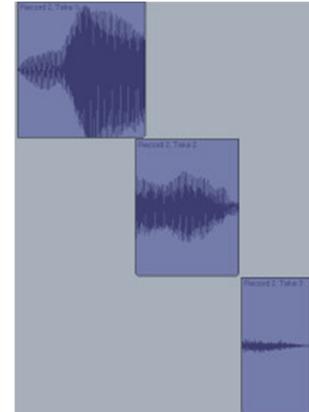
Most of the editing techniques we've discussed to this point are relatively straightforward and easily accomplished, even if they may get tedious at times. Hopefully those techniques will cover us for 90% or more of the editing we'll need to do for any given vocal project. In fact, were it not for the fact that I enjoy an intellectual challenge from time to time, and that the more frustrating such a challenge seems on the front end, the higher the level of satisfaction if/when a solution is found, I might even say I'd hope the techniques we'd already discussed would cover 99.9% of any given vocal project. To put it quite frankly, the territory we are about to enter, which I will affectionately call "microsurgery" since it involves working at an extremely detailed level, sometimes even within a single syllable of a word, can be extremely challenging and extremely frustrating. Moreover, there are no guarantees of success, even after all our time and frustrations. Nevertheless, sometimes these techniques can make the a significant difference in our end results, and even get us out of that precarious position that lies between a rock and a hard place.

The general situation that would lead to the use of these techniques is when you have no single take of a phrase that works it its entirety. Nor are there any portions of different takes of that phrase bounded solely by convenient split points that can be used to create a composite take that will do the trick. Or perhaps, in a somewhat rosier picture, you do have an entire take, or easily splittable portions of takes, that could create an acceptable composite take. Still, there are pieces in different takes, without convenient split points, that could create an even better composite result -- if, that is, you can work some magic and combine things at a non-optimal point of juncture. Let me give an example:

The song you're recording ends on a long held out note on the ever-popular word "love". You have three takes of the word, which is its own phrase in this song. The first take starts out strongly, but somewhere just after the middle of the word, the singer's breath runs out and the note gets weak and extremely out of tune -- beyond the point any tuning software can rescue. The second take has a bloop on the front end of the word, but ends just perfectly, with one exception: It dropped the final "V" sound on "love", so it sounds more like "luh". The third take is pretty nasty

all around, but does at least get one part done well: it has a very nicely pronounced "V" on the end of the word. Unfortunately, there is no chance to get the singer back into the studio because he is off on a month-long honeymoon safari in Zimbabwe, and the album mixes are due prior to his return. It's okay, you can say it: "This really sucks!" Yep, it sure does.

You have probably already guessed what we will be forced to do if we aren't satisfied with using one of these inferior takes as our "best we could do with the materials at hand" compromise. That's right, we will need to find some way to combine the start of the first take with the end of the second take, then add the "V" sound from the third take at the end. That may sound simple, but the problem is we're going to need to split a held out vowel sound to try and splice both ends of that held out sound together, without generating noticeable artifacts. There probably also isn't a good split point to get at just the "V" sound, either, but that isn't quite as bad as trying to split a vowel in the middle. Taken in their native state, just mixing the applicable portions of the three takes together might sound something like this:



"Love" - Badly spliced combination of three takes.

Here is what that result looks like when mixed:



Notice especially the relatively harsh waveform where the first two takes overlapped. The nastiness that created was quite obvious in the audio result. Let's consider some of the problems faced in this example:

- Volume differences: The first and second takes had noticeable volume differences between them, which will make it hard to create a smooth transition.
- Waveform synchronization issues: The wave forms are highly unlikely to be overlapping nicely in their native state, making it tough to make them blend without creating glitches near the transition points.
- Length problems: I mentioned a held note, but this combined sample is shorter than might be optimal.
- Possible minor tuning issues: While these seem to be reasonably in tune with each other, there may be slight variations that make it difficult to create a smooth transition.
- Possible timbral differences: It is quite possible that, along with the volume differences, there may be slight differences in timbre between takes that could also make it tough to create a smooth transition.

We won't know whether we encounter these issues, and just how bad they are, until we try to overcome whatever is causing the audible problems we heard. However, it is a sure bet this change will take a bit of finagling, and there are no guarantees we'll be able to arrive at a good solution. Ah, but it's a challenge!

For the moment, I won't worry too much about the volume differences between clips, because it will be hard to tell how bad things are in that department before getting the waveforms in better sync at the transition points. SONAR 4 provides the ability to see one waveform through another you are dragging on top of it. If we zoom in far enough to see the individual wave cycles, we can try and manually line up peaks and troughs and zero crossings. It isn't quite as simple as that, though -- not that lining those points up is simple since the tuning considerations, among other things, could create differences in wave cycles. The harmonic structure of the waves will be changing over time for something as rich as the human voice can be, and we need to try and match things up at the transition point as closely as possible to avoid rude artifacts around the transition points.



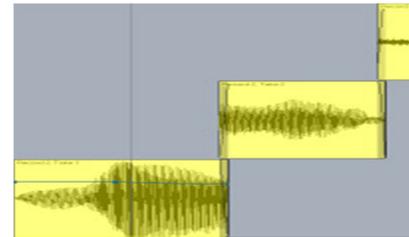
Overlapping the first two takes, we can synchronize the waveforms as shown at left. It looks like these two takes are pretty well in tune at the transition points, as they line up nicely. Their amplitudes and harmonic structures don't match up as nicely. The leftmost clip is louder, and a bit less complex in harmonics. Nevertheless, synchronizing the waveforms at both splice points, then using the overlap cropping tool with automatic crossfading turned on, and finally adjusting the crop points to ensure a zero crossing is hit where a new clip starts and an old clip ends, results in a much improved composite:

["Love" - Better splice aligning waveforms and using crossfading at overlap points.](#)

Here is what the composite waveform looks like:



Note that the squared off area near the transition point is no longer there. There is still a decided drop in level between parts of the word, though, so perhaps using clip envelopes will help. Experimenting with this a bit more, I also found that changing the crossfade to linear on both ends, rather than the default of fast fade-in and slow fade-out, also helped smooth things further. Have a listen:



["Love" - Even better splice utilizing clip gain envelope and linear crossfade.](#)

Here is a look at the composite waveform:



There is still a bit of a bump at the transition, but this is much smoother, and we may just have saved our singer from an early divorce.

Let's recap what we've done thus far: First, we sliced up a single word into various component parts so we could borrow acceptable parts from different takes. Next we aligned the wave cycles of the two clips in the region where they were overlapping to avoid nastiness around the splice points. Finally, we used a clip gain envelope to try and create more consistent volumes near the splice point, and tried different types of crossfade curves to optimize the smoothness of the transition, ending up with a linear crossfade in this case. This isn't a job I'd wish on my worst enemy in most cases. Amazingly enough, though, it didn't come out half bad, and I would venture that an end listener would be unlikely to notice this wasn't a word sung in one pass, especially when the vocal is put in the context of the instrumental tracks.

Of course, that's not all that can be done. We mentioned the possibility the note may not have been held out long enough. There are several ways to address this. Perhaps the easiest way, and the one which will likely bring the best results in many cases, is to use SONAR 4 Producer Edition's MPEX time and pitch stretching algorithm as embedded in the Time/Pitch Stretch 2 plug-in. The MPEX algorithm allows very high quality time stretching, and if you have a single word, holding a single note, this may well result in the most natural result. In the case of our example, we could simply bounce our results thus far to a clip, select the clip, call up Audio Effects/Cakewalk Time/Pitch Stretch 2, and then pick our settings to choose the MPEX algorithm, the best quality stretch for a voice or single instrument, and the degree of stretching we need. If there were multiple notes in the word, though, unless

the durations of all notes needed to be stretched by the same percentage, this method would not have been very helpful.

Another option would be to stretch all the individual components of the composite word before combining them, also with the MPEX algorithm. This allows the flexibility of stretching different components differently, in which case it would be best to align the wave cycles after the time stretching operation as the alignment might change. Still another option, which might be necessary if you don't have SONAR 4 Producer Edition (MPEX is not available in SONAR Studio Edition), is to simply cut up the words, further, copy an extra piece with only the vowel one or more times, then repeat the type of aligning and crossfading we did above as many times as needed to get things long enough. This kind of operation can be extremely difficult to get right due to the potential for making things sound like a loop or having very strange harmonic and dynamic things going on (e.g. due to a natural change in harmonics and/or dynamics over the course of a held note). However, in a pinch, it might be better than nothing, though not a whole lot better.

I purposely gave what I consider to be a nasty example above to try to demonstrate how even a nasty problem like trying to splice together a held out note can sometimes be resolved acceptably if we use a bit of ingenuity, experiment a bit, and have a lot of patience. However, where these techniques tend to be more useful is when dealing with phrasing issues, where shuffling things around in time creates some small gaps or too tight spaces. You can cut things up at less-than-ideal places, as long as you are willing to do a bit more work to make the splices work. Time stretching (or contraction) can be used to try and avoid some of the unnatural splicing. Longer or shorter crossfades, with various fade-out/fade-in curve combinations, can be used to try and smooth over lumpy overlaps. Individual syllables can be pitch shifted or tuned. Clip envelopes can be used to deal with more volume leveling considerations within or between clips.

These ideas are by no means comprehensive. The bottom line is to determine what the problem is, and, via your familiarity with the various tools and techniques available to you, try and find some way to resolve the problem and make what may have been put together in a most unnatural way sound natural. Of course, sometimes it is best to just have the singer redo the part, if that is an option, and a general rule of thumb is the better the singer, the less frequently you are likely to need to resort to microsurgery and Frankenstein phrases.

Background Vocal and Doubling Considerations

While most of what we have talked about to this point has been discussed primarily in the context of lead vocals, it also applies equally to lead vocal doubles, background vocals, and background vocal doubles. Background vocals and vocal doubles bring some special considerations, though. For example, generally the lead vocal will be paramount, and a double of the lead vocal and any background vocals need to support and synchronize with the lead vocal. A lead vocalist's timing slop may define the song, but a background vocalist's timing slop will often detract from it. At the same time, vocal doubles and background vocals are generally mixed far less prominently than the lead vocal. Thus, things you could not get away with in the lead vocal may be just fine in background vocals, especially as part of a thick background vocal blend.

In no particular order, here are a handful of practical tips that may help in adapting the techniques we've already discussed, while adding a few more, for background and doubled vocals:

When it comes to tuning background vocals, the thought of hand tuning, say, doubled five part harmonies is pretty daunting. The good news, especially when you are doubling background vocals, is you can get away with a lot looser tuning than with a lead vocal. In fact, slight differences in tuning between doubled vocals are part of what can make background vocals sound rich. As such, I rarely do graphical tuning of background vocals. Instead, I tend to use automatic mode tuning, but on a phrase-by-phrase basis, being careful to set the automatic tuning mode parameters to only include notes that are relevant to the phrase in question, thus minimizing mistunings of too-far-off notes. Additionally, you can select all takes of a single harmony part for a given phrase, then apply destructive tuning to them in one go, thus letting you use the actual tuned phrases in your comping decisions. Any tuning errors that end up in the result may make you more likely to need to split phrases more finely with background vocals than

you do with lead vocals (i.e. to chop out tuning errors in an otherwise good take). However, this may be more likely anyway for reasons we'll discuss in a moment. And that moment is now...

The need to have background vocals and doubled vocals be tight with each other, and possibly with the lead vocal, often results in the need for increased microsurgery when dealing with background vocals. I would still try and avoid cutting up syllables in the middle wherever possible. Still, the good news is that, because background vocals will be mixed less prominently in most cases, you don't have to be quite as careful to avoid minor artifacts in dealing with this microsurgery as you do when dealing with lead vocals. Something like a minor level blip may not be noticeable in the blend.



One trick for tightening background vocals against a lead vocal, or against a reference background vocal if the lead vocal has different rhythms than the background vocal, is to run each background vocal through a gate, but control the gate's opening and closing off the lead vocal or reference background vocal. In the hardware world, this is typically done using a sidechain input into a gate, where the sidechain is the lead vocal and the main signal being compressed is the background vocal. In the software world, one challenge is that the lead vocal and background vocals will be on different tracks, and most plug-ins do not allow using one track as a sidechain input and another track as the main input. One exception, which happens to work

nicely for this particular application, is [db audioware's dB-D dynamics processor](#). The dB-D plug-in allows you to link separate instances of the plug-in through send and receive channels. Thus, in this case, you would set up one instance of dB-D on your lead vocal, setting its send channel to, say, "A". Then you would set up one instance of dB-D for each background vocal you wanted to gate off the lead vocal, setting the receive channel of each instance to "A", also. This will allow controlling the opening and closing of dB-D, which can serve as a compressor or gate, on the background vocals via the lead vocal. Of course, there is nothing to say that the background vocals will be tight enough that gating them off the lead vocal will sound natural, as it could mean cutting in at out at inopportune times. However, if the background vocals are already tight, be it naturally or after editing, this technique can make them even tighter, and the technique seems to work pretty well in practice. Here is a "before" sample of a lead vocal, doubled lead vocal, and two double harmony vocals:

["Lust" - Raw background vocals.](#)

Now here is the same section with the lead vocal double and background vocals gated off the lead vocal:

["Lust" - Background vocals gated off lead vocal.](#)

The effect is somewhat subtle, but the second example should seem slightly tighter than the first example.

One "trick" that has been used for decades now is doubling vocals to make them sound fatter or richer. The idea is that subtle differences in tuning, timing, and timbre between one performance of a part and the same voice performing the same part a second time combine to make the sound fatter. Typically, the double of the vocal is mixed lower than the main vocal, possibly similar to treatment of the background vocals, depending on how much of the effect is desired. One further trick that can help make things even fatter is to clone the double, then nudge one of the clones a little bit forward in time and the other the same amount backwards in time, then pan one of the doubles all the way left and the other all the way right. Here is a lead vocal before doubling:

["Lust" - Lead vocal before doubling.](#)

Now here is the same section with a doubled lead vocal to which the cloning/nudging/panning trick has been applied. The doubles have been mixed 6dB down from the main vocal. Also, the lead vocal uses a harmonic exciter to bring it out slightly in the mix, while the doubles do not:

["Lust" - Lead vocal after doubling.](#)

You may have noticed that, while the doubling makes the result richer here, it also takes away some of the intimacy of the original vocal. Whether that is a good thing or a bad thing will depend on the type of song and the singer. As a general rule, I am more likely to use doubled lead vocals on pop and rock songs, and less likely to use them on country or acoustic songs. Similarly, I am more likely to use a doubled vocal on an up-tempo song than on a ballad.

Of course, even when you do double a vocal, you don't necessarily have to do it throughout the entire song. While I only rarely double lead vocals in verses, I fairly commonly double them in choruses, and slightly less commonly in bridges. The idea is to go for maximum intimacy on the verses, but then to make the voice richer when the section with the hook comes in. This is partly just to reinforce the hook and chorus in general, and partly to help the lead vocal maintain power against any background vocals, which are more likely to be present at that point. With respect to bridges, and sometimes pre-choruses, it depends on the function of the section and the vocal arrangement. For example, if the section needs to build power over the verses, doubling can help.

While I don't always double lead vocals, I almost always double background vocals. Doubling tends to make a vocal more diffuse than an undoubled vocal, and background vocals tend to benefit from being more diffuse -- it helps them sit better behind the lead vocal, which should be more prominent and seemingly more "out front". An exception might be in cases where it is desirable to have the background voices stand out as being unique, and on a relatively similar level to the lead vocal. For example, if I were going for a live bluegrass band feel, especially with some type of slow, mournful song, I'd be inclined to try undoubled background vocals.

While doubling a single voice for background vocals can help make the vocals more diffuse and add to the richness of the parts, it is usually possible to achieve even richer results by using multiple voices (i.e. from different singers). There the differences in timbre between voices add to the overall effect. One technique is to have each singer sing each of the harmony parts, rather than have each singer double only a single part. Of course, you can also double each singer to combine doubled vocals from each singer with doubled vocals from additional singer(s). There is no need to stop at two singers, either, though the more vocals you get the harder it is to keep things tight and avoid sounding like a choir.

What if you're a solo operation, with no access to other singers, but you'd like to try the notion of mixing additional voices with yours for doubling background vocals? One possibility is to use a plug-in that can do formant shifting, sometimes known as "gender bending", to alter some of the doubled vocals. One plug-in I have used for this purpose, with somewhat mixed results, is [RBC Voice Tweaker](#). The problem hasn't been so much how artificial the gender-bended, or just slightly altered, results can be -- tasteful use of the formant tuning parameters can help keep things within reason. Rather it was in the potential for also getting nasty glitch-type artifacts in some cases. Another possibility is using a vocal synthesizer, such as those based on Yamaha's Vocaloid technology like [Zero-G's MIRIAM](#), LOLA, and LEON. While Vocaloid voices can be tough to make sound natural in exposed settings, such as for simulating a lead vocal, using them for blending with real background vocals can add richness to the background vocals without exposing some of the less natural-sounding aspects of the synthesized voices. (If you'd like to hear blended human and LOLA voices in action, check out "Undertow" on [my Soundclick.com page](#).)

Doubling vocals can be a pain, especially for lead vocals. You need to get as near identical phrasing as possible, keep diction in a similar ballpark, keep vocal inflections and ornaments very nearly identical, and so on. That can be tough to do. Various techniques, generally referred to via the umbrella term "artificial double tracking" (ADT), can be used to try and gain similar benefits to real doubletracking without the pain and potential pitfalls. The idea is to track the vocal once, then use some kind of manipulation to simulate having tracked the voice more than once. For example, you can use delay to alter the timing of a cloned copy of the main vocal (or route the main vocal to a pre-fader bus, then treat that bus like a second vocal track), or just nudge the cloned copy backward or forward in time, then use other effects to alter aspects of the vocal beyond just timing. Here is an attempt at ADT for the example given up above for the real doubling case:

["Lust" - Lead vocal with ADT.](#)

Instead of real doubling, I routed the main vocal to a pre-send bus, then inserted the Sonitus:fx delay with a 6ms 100% wet setting and 40% feedback, 10% crossfeed, plus a little high-end rolloff. I followed that up with Antares

AutoTune 4, set to a chromatic scale with -10 cents of detuning, then the Sonitus:fx Modulator with a relatively light ensemble setting. The artificial double was then mixed about 6dB down from the main vocal. While I prefer the natural double, this one sure required a lot less work!

When working with doubled background vocal parts, using a single voice, after doing enough extra takes to feel confident I've got enough raw materials for two of each part, I find it useful to use SONAR 4's multi-lane tracks to comp, and possibly even tune and edit, both performances of a given part on the same track. Instead of looking for the best take for each part, you are looking for the best two takes. However, you are also looking for two takes that blend together well, and also blend well against the lead vocal. Thus, you may want to use the Isolate capability of the Mute tool to narrow down the choices of takes, but then you'll need to unmute additional takes to hear them auditioned against each other to try out various combinations. Since all takes would have, optimally, been sung against the lead vocal, they should be reasonably in time with each other. This same approach doesn't work as well for doubling lead vocals, because there wouldn't be the final lead vocal to reference when doing any of the takes. However, if you are doing multiple real doubles of lead vocals, and are tracking those after the final lead vocal has been comped, the same technique can work. Once you have your two chosen takes for comping, you can move one set to a new track so it can be panned and mixed separately.

Submixing Background Vocals

While background vocals will necessarily be composed of individual voices and parts, once you get a balance and blend of the voices working, it can often be helpful to work with the background vocals as a unit. Prior to SONAR 3's bus structure enhancements, doing this generally required bouncing to create a new submix track, then muting the original background vocals and using the submix for any subsequent processing. If you changed your mind on anything related to the blend, you'd have to go back and adjust that, then bounce again, and replace the original submix with the new one. One of the tricky things was that you couldn't hear the submix-level processing while working on getting the blend in the submix, so there could be a fair bit of trial and error involved en route to your desired result.

SONAR 3 added submix buses, so it is now easy to route all the background vocals to a submix bus, do any further processing at the bus level, then route that submix bus to your main mix. Of course, that is assuming your CPU is sufficiently powerful to handle whatever else you have going on in the mix, but the SONAR 4 Freeze function provides the ability to free up CPU in other areas while you're working on getting that perfect background vocal blend. Additionally, once you're sure you've got the track level insert effects the way you want them, SONAR 4 will let you freeze the individual background vocal tracks to free up whatever CPU their effects were using. If you're talking about a dense set of doubled background vocals, even having a few insert effects on each could eat up a fair amount of CPU power, especially on an older processor (like my Athlon XP 1600+).

While they don't affect the aural side of background vocals, SONAR 4's new Track Folders do aid in thinking about background vocals as a unit, and doing some editing operations on them as a unit. For example, you can put all background vocals in a track folder for organizational purposes, then collapse the folder to regain space and operate on the background vocals as a group for operations such as muting and soloing them. You can also use the composite clips in the track folder's folder track to do things like copy a whole section of background vocals from one part of the song to another (e.g. for using the same set of background vocals for multiple choruses). Thus, the combination of track folders and submix buses goes a long way toward letting you operate on the background vocal submix as a unit while still preserving the flexibility to get at the individual tracks.



We'll talk about vocal effects in general shortly, and the individual background vocals will likely each have their own set of effects, such as compressors, equalizers, and so on. What those effects are, and whether they are consistent from voice to voice will depend on your intentions and the voices at your disposal, though, for blend purposes, odds are you'll process them relatively similarly to each other.

Effects used at the individual vocal level are geared primarily toward getting a good submix of the background vocals. By contrast, effects used at the submix bus level are geared toward fitting the background vocals, as a unit, in with the rest of the mix. For example, compressors used at the individual vocal level may help smooth out the dynamic peaks and troughs of an individual voice, for example to compensate for mic technique issues. Compressors used at the submix level will help smooth out the dynamics of the entire background vocal submix.

My general tendency in differentiating with individual vocal level effects and submix level effects is to hold off as much as I can for the submix level, to allow treating the background vocals as a single mix unit. However, if I want different treatments on individual voices -- for example, to impart varied timbral characteristics on the same voice when doubling parts, this has to be done at the individual voice level. Other control that will need to be asserted at the individual voice level includes separate panning of voices to establish the individual voice positions within the group, any vocal blend-oriented gain settings and level riding, and any arrangement-oriented automated muting. Then the submix bus level will be used for adding group-level compression, EQ to fit the submixed background vocals in the mix, any delay and/or reverb sends, and any special effects (e.g. filters, stereo widening, etc.) intended to apply to the group vocals.

Vocal Processing

Dolly Parton is reputed to have said, "it costs a lot of money to look this cheap." In the world of vocal production, an apt paraphrase might be, "it takes a lot of work and processing to sound this natural." While a singer might sound just great singing acoustically in a room, when you stick a mic in front of her, put a preamp in back of the mic, use A/D converters to convert the signal to digital form, capture that digital representation on hard disk, then mix the voice in with various recorded instruments and additional voices, what came across via the airwaves and the acoustics of the room may not be recognizable in the recording. Issues such as the limitations of the equipment used, especially the microphone's pickup pattern and response characteristics and recording methods for getting down the other instrumentation and background vocals, among others, can mean that what sounds completely natural in the room sounds anything but when played back through your speakers in a mix. Furthermore, various musical genres carry with them certain listener expectations for production styles that include vocal treatments that may or may not be geared at keeping the voice sounding natural.

In this section of the article, we will give an overview of some of the common processors and effects used in vocal production. While the list will by no means be comprehensive, the idea is to make a basic survey of the field, indicating the basic needs that often suggest using a given type of processor, providing any tips that might spring to mind, and mentioning at least a few of the products that are out there to provide that functionality. Wherever applicable, I will try and use plug-ins already included with SONAR 4 Studio Edition or SONAR 4 Producer Edition as examples, though I will also mention a few of the third party offerings. My apologies to any vendors who are not mentioned here. This list is not meant to be comprehensive, and the products I mention will be limited to ones with which I have had personal experience, preferably positive experience, wherever possible.

It might be typical at this point to begin with the most commonly used processors and effects, for example compressors/limiters, equalizers, delay, and reverb, then move toward less commonly used processors, eventually getting to the really esoteric stuff. However, I'm going to take a slightly different approach. In particular, while there is no set order in which processors and effects must, or even should, appear in a vocal chain, I am going to use the direction of signal flow in one possible vocal chain as a starting point.



When we look at a vocal chain that starts with the singer, the first element in the chain after the singer is generally the mic. Hopefully we have a mic with characteristics that are well matched to the singer's voice and the context of the song. However, what if we don't? What if we have a highly limited mic closet, and none of the mics we've got at hand flatters the singer's voice, and perhaps even exaggerate less desirable characteristics of the singer's voice?

Enter Microphone Modeler from Antares. The idea behind Microphone Modeler is that all microphones have individual characteristics. If those characteristics can be analyzed in terms of

their effect on audio put through them, then it should be possible to subtract the characteristics of the microphone which was actually used to record the singer, for example my The CAD Equitek E-200 I use (but which does not flatter my voice), and add on the characteristics of another mic -- perhaps a Neumann U47 or AKG C12A, for example. Mind you, I don't have any Neumann or AKG mics to tell you whether my voice through an E-200 which is "transformed into a virtual C12A" by Microphone Modeler actually sounds like my voice would through a C12A. What I can tell you, though, is my voice sounds much better through that "virtual C12A", and, for that matter, through a number of other modeled microphones. Some of those modeled microphones, including the C12A, sound much closer to what I hear in my head than does the E-200 on its own. The aural differences may be subtle, but the difference in listeners' impressions, including my own, has not been. I might add that, prior to finding Microphone Modeler, I had tried using EQ to achieve similar results, but nothing I'd tried had provided similar results. As a result, Microphone Modeler tends to find its way onto all of my personal vocal tracks. I've also found that even singers who sound great with the E-200 may sound even better with another virtual mic model. Thus, experimenting with different models has become a standard part of my post-tracking process. Essentially, it is like trying to match the mic to the singer, but after the fact.

In addition to modeling the microphone itself, Microphone Modeler models important controls many microphones. Those controls include whether a low cut filter is engaged or not (and at which setting for mics with multiple settings) and which pickup pattern is used for mics with multiple patterns. Additionally, for mics for which proximity effect is applicable (e.g. mics with cardioid pickup patterns), you can change the singer's "virtual proximity" to the mic. It also models a tube preamp with varying degrees of drive. While there are other plug-ins out there that do impulse response modeling which can be used to graft the characteristics of an audio device onto an audio signal (e.g. Sony's Acoustic Mirror and the SIR reverb), I am not aware of any other than Microphone Modeler which first attempt to remove the characteristics of a similar device from the audio signal.

Speaking of preamps, they are next up in the vocal signal chain. While you will be using some kind of preamp in the process of recording, some preamps are used as much for the flavors they impart as for their signal boost, and perhaps you are not recording through one of those but are looking for a similar effect. A variety of plug-ins cover this territory. If the flavor you are looking for is an extreme color, such as you might get by running a vocal through a guitar amp, SONAR includes both AmpSim and ReValver SE. However, it does not provide any plug-ins for the more likely scenario of simply trying to simulate some "tube warmth". Some plug-ins I have successfully used in this area include Antares' Tube (basically like Mic Modeler without the mic modeling part, and with some additional enhancements including a second tube preamp model), PSP Audioware's PSP MixSaturator, and SilverSpike's freeware RubyTube. Each of these plug-ins has its own specific controls and characteristics. At the most basic level, though, you simply turn up a drive knob (or slider) until you get to the point where you are feeling the desired level of "tube saturation".



We'll be getting to some of the most commonly used vocal processors shortly. Before we do, however, this is a good time to mention a historical note. In the digital recording world of 24-bit DAWs, it is common to try and keep vocal recording chains ahead of the recording medium as short as possible, preserving maximum flexibility for later on, such as at mix time. Thus, if we are look at our processing chain as a model of a physical world processing chain, this could be the point at which the recording medium would enter the picture, with anything after this point being effects added at mix time. However, not so long ago, when the norm was to track everything, including vocals, to analog tape, working within, or around, the limitations of analog tape often resulted in putting at least some of the processing elements in the overall signal chain ahead of the recording medium. For studios with enough hardware processors, such as compressors, a common practice would be to track with light processing -- for example, just enough compression and/or limiting to get a good signal on tape -- then do additional processing come mix time. However, studios with more limited hardware might be forced to track as much to tape as possible, for example, because they didn't have enough compressors to go around at mix time. The native plug-in world tends to make limitations on the number of processors a thing of the past, at least if you have enough CPU power and/or use Freeze to get around any CPU power limitations. I mention this historical point because, if you are trying to model the vocal chain used in some historical context, and are using a tape simulator to model the tape's contribution to the processing chain, where you put that tape simulator may depend on what you assume in terms of the facilities available in the virtual studio corresponding to your modeled historical chain.



There are also a number of plug-ins that work to simulate recording to tape, including Cakewalk's own FX2 TapeSim, which is included with SONAR, PSP VintageWarmer, and the above-mentioned PSP MixSaturator. While I'm not a big fan of tape simulators for vocals (I do like them for mixdowns), this is yet another flavor that can be brought into play if desired, and it might be useful if you are specifically trying to simulate some traditional vocal recording chain.

Perhaps the most commonly used vocal processors are compressors and limiters. Both help even out dynamic levels by reducing the level of a signal that exceeds a certain threshold, be that a fixed threshold defined by the processor or a user-controllable threshold. The idea is that, once the level of the vocal exceeds the threshold, it will be automatically turned down by the processor. How much it is turned down depends on the processor's ratio. For example, if the threshold is set at -10dB and the vocal signal hits -7dB, thus exceeding the threshold by 3dB, a 3:1 ratio on the compressor would result in the signal's being turned down to -9dB -- i.e. only 1dB over the threshold, which is one-third of the amount by which it was originally over the threshold. The difference between a compressor and limiter is generally related to the ratio used. Most anything over about a 10:1 ratio is considered limiting, while lower ratios (i.e. above 1:1) are considered compression. While a compressor or limiter only turns signal down, most compressors and limiters have the ability to apply makeup gain to turn the signal back up. The result is generally an apparent increase in average loudness. This is because you will typically be turning up the compressed or limited highs to the level where the old, uncompressed highs were, but the lows, which were likely not affected by the compression, are being turned up, too. This makes for decreased dynamic range, and a signal that is louder on average.



SONAR 4 Producer Edition includes the excellent Sonitus:fx Compressor, which is technically a compressor/limiter, as well as an older Cakewalk FX1 series of dynamic processors which are also available in SONAR 4 Studio Edition. (Note: The FX1 series are not installed during the SONAR 4 installation by default. However, they can be checked for optional installation during the SONAR 4 installation, or installed separately later.) SONAR 4 Producer Edition also includes Sonitus:fx Multiband, a multi-band compressor/limiter which can process different ranges of the frequency spectrum as if it were using different compressors on each band. This can be useful, for example, for doing de-essing (see below) while also compressing and limiting the vocal. SONAR 2 XL users will also have Timeworks' CompressorX, though it is no longer included in the SONAR 4 packaging, so newer SONAR users would not have it unless they've purchased it separately. Other third party compressor/limiters I've found useful include PSP MixPressor, Waves RComp and RVox (a special compressor/limiter/gate designed specifically for processing voice), and the previously mentioned dB audioware dB-D. While all compressors operate under similar principals, each has its own sound. Thus, it is worth experimenting with any compressor plug-ins you have available to see what works for a given voice and song.

Here are a few tips for using compressors: First, there is no need to limit yourself to a single compressor in the signal chain. It can sometimes be useful to use multiple compressors. For example, the first compressor in line can take care of any stray peaks, probably working more as a limiter than a compressor, then the next one can work on a more representative signal, not having to worry about a few aberrant peaks here and there. Secondly, the position of the compressor(s) in your signal chain matters. For example, compressing a signal after EQ is added can bring very different results than compressing that same signal before the same EQ is added. The reason is because most processors, including EQ, affect the dynamic levels of the signal, and sometimes the effects of those changes can push the signal over the threshold where it would not have crossed the threshold before. Also, even a signal that would have crossed the threshold before will be reduced differently if EQ is added before the compressor since the compressor also reduces any boost from the EQ. Whereas the compressor won't affect EQ results when the EQ is after the compressor. The same is true of other effects and their order a vocal chain including compressor(s). There is no right or wrong order. It just depends what you are trying to achieve.

One special type of compressor is called a de-esser. As its name might suggest, it is used to reduce the level of "S" sounds and other sibilance (e.g. "sh", "ch", etc.). If your singer's esses sound like a cat hissing, this kind of processor

may be for you. The way this works is the compressor circuit in a de-esser only kicks in when there is a certain amount of energy in a specified frequency range, which is adjustable in most de-essers since the frequency range where sibilance exists will differ between different types of voices. It is like having a side-chain that only includes the signal within the specified band of frequencies. What actually gets compressed when the compression circuit is triggered depends on the de-esser, and possibly the mode of operation set within it. For example, the whole signal may be compressed, even though only a specific band is triggering the compression. Or a certain range of frequencies, which may or may not be identical to the range triggering the compression, may be compressed. Or anything above a specified frequency may be compressed.



Though SONAR doesn't actually include a dedicated de-esser, SONAR 4 Producer Edition does include two plug-ins that can provide some sibilance reduction. Sonitus:fx Multiband can work as a narrow band de-esser, where the range of frequencies compressed will be the same as the range of frequencies that triggers the compression. Other bands of Multiband can be set to either pass the audio straight through or to serve as more classic compressors and/or limiters. Sonitus:fx Gate can also reduce sibilance in its ducking mode. Whereas a typical gate or expander is used to reduce low-level signals further, for example to reduce or eliminate background noise,

Sonitus:fx Gate also provides a ducking mode that can be used to reduce signals above the set threshold according to a depth control. Since the portion of the input signal used to trigger the ducking action can be constricted based on high and low cut filters, the range between these filters can be set to be the range of sibilance, and ducking can occur only when the signal rises above the threshold in that range. Third party, dedicated de-essers include the powerful Waves RDeEsser (Renaissance DeEsser) and the very economical, yet still highly functional, dB audioware dB-S.

One final type of dynamics processor that can sometimes be useful for vocal processing is the noise gate or expander. Unlike a compressor, which turns a signal down when it crosses above a certain threshold, noise gates and expanders turn a signal down or off when it crosses below a certain threshold. The term "expander" comes from its expanding the dynamic range -- i.e. making the low parts of the range lower by turning them down further, while leaving the high parts of the range untouched. The difference between a gate and an expander is the gate turns the signal off altogether, whereas the expander may just turn it down somewhat. Either can be used for reducing noise in quiet parts of a track. The gate goes for total silence, while the expander typically goes for reducing background noise without getting to the sometimes unnatural effect of total silence. While it was pretty common to use expanders or gates for treating vocals to lessen or remove the effect of room noise in the analog realm, comping and slip editing can lessen or remove the need for these devices. Noisy areas can simply be edited out, or clip envelopes can be used to reduce noise levels in ways that provide much more control than a gate or expander does, without the potential ill effects of, for example, chopping off an under-enunciated consonant. However, when time is of the essence, and background noise in exposed passages is a problem, these processors can still be useful. SONAR Producer provides the excellent Sonitus:fx Gate, and all versions of SONAR provide the FX Expander/Gate (though its installation is optional). Both plug-ins can be used as expanders or gates. Dedicated third party gates are tough to find, but some compressor/limiters do incorporate expander/gate functionality as well. For example, Waves' RVox and RChannel, along with Wave Arts' TrackPlug, come to mind. RVox is a specialized voice compressor, while RChannel and TrackPlug are "channel strip" processors incorporating dynamics processing and EQ.



Equalizers, or EQs for short, are, along with compressors, delays, and reverbs, among the most common vocal processors. I have to tell you, though, while I do use EQ on vocals often enough, I prefer not to, especially with lead vocals. In simplistic terms, an EQ is more or less a set of frequency-dependent volume controls, where you can raise the level of the signal at some frequencies, while leaving other frequencies alone, and lowering the level of still other frequencies. In practice, it isn't quite that simple as the various bands of an EQ will interact with one another, but the point is that an EQ is used to alter the tonal balance of the signal it processes. From a purely philosophical standpoint, because the vocal is so important to most recordings, I prefer to let the tonal balance of the voice come through. I'd prefer to move other things out of the way if necessary, rather than to change the tonal balance of the voice to cut through the mix. In practice, though, it isn't always so easy to do that, and there are always those times

where it is desirable to alter the tonal balance of a specific singer's voice, at least in the context of the musical style and song, too. Also, sometimes EQ can be helpful for lessening or eliminating annoyances like popped "P" sounds, certain types of low frequency mic stand noises, electrical hum, and more.



While there are some home-stereo style graphic EQs available in plug-in format, where the frequencies and bandwidths of each band of EQ are fixed, most plug-in EQs are of the parametric or semi-parametric type. With parametric equalizers, you can adjust the center (or end) frequency and bandwidth (or slope) of each band of EQ. A semi-parametric EQ is like a parametric EQ, except there may be some decisions that are made for you, such having a fixed bandwidth (but sweepable center frequency). SONAR Producer Edition includes the fully parametric Sonitus:fx Equalizer. All versions of SONAR include the historic Cakewalk Graphic EQ and Parametric EQ, as well as the DSP-FX-based FxEq. SONAR 2 XL also included the Timeworks Equalizer. One particularly nice thing about the

Timeworks Equalizer was it had an FFT-analyzer built into it, so you could see the frequencies of the incoming (or outgoing) signal while setting your EQ curve visually against the FFT graph. Some third party EQs I particularly like include PSP Audioware's PSP MasterQ and Waves' REQ. Like compressors, equalizers can often have a flavor, so it is worth trying different plug-ins. Because lead vocals tend to be so prominent in a mix, one rule of thumb might be to use the highest quality EQ you have on the lead vocal, even if that means consuming more CPU cycles than a lesser model.

Somewhat related to the equalizers are exciters and enhancers. Exciters typically manipulate or add high-end harmonics to a signal to psychoacoustically manipulate our impression of the signal's brightness and presence without adding high-end EQ. Other forms of enhancers manipulate phase information to monkey with our perception of how close a signal is. While I'm not going to get into more detail on the principles of operation in these types of processors (I'm not sure I could if I wanted to), the key is that they can be used to manipulate the presence of the vocal in the mix. For example, if a vocal is getting buried in a mix, and gain changes and EQ are not the right answer, using a harmonic exciter may make the vocal seem more present. This may let it stand out in the mix compared to other signals, which then, by comparison, seem further back. SONAR does not include any exciters or enhancers, though Cakewalk's Project5 does include the HF Exciter (HF for "high frequency"). Some third party harmonic exciters include the DSP-FX Aural Activator (my personal favorite for lead vocals) and PSP Audioware's PSP MixTreble. The BBE Sonic Maximizer is an enhancer of the phase manipulation variety.



Thus far, the processors we have discussed are primarily focusing on the direct sound of the vocal. In a real room, though, there is also indirect sound, for example from the voice's bouncing off walls, ceilings, furniture, and so on. There are two main categories of effects related to indirect sound: delay (sometimes called "echo") and reverb. These primarily affect our impression of the space surrounding the voice, rather than the voice itself, though they can also have bearing upon our overall impression of the voice.

Delays provide discrete reflections of the sound, of the sort one imagines coming back after shouting into a canyon -- "HELLO... Hello... hello..." In the real world, the idea would be like that canyon echo, where you hear the same, recognizable words coming back to you, with a lower volume each time, until they ultimately fade into nothingness. However, in the vocal processing world, most delay plug-ins won't constrict you to real world simulations, and you can often have pretty detailed control over how many echoes there will be, how their volume will differ from the "tap" before, and so on. Speaking of taps -- a "tap" is a controllable instance of an echo -- many delays allow configuring multiple taps, for example with their own timing parameters (polyrhythms, anyone?). Even single tap delays can provide multiple echoes, though, via controls which are generally labeled "feedback", which feed a portion of the delayed signal back to the input of the delay (i.e. a "feedback loop", and, thus, the name of the parameter), thus keeping the effect going beyond the initial tap. In music productions, delays are often used for rhythmic effects, rather than for simulating naturally occurring echoes. For example, a sixteenth note or quarter note delay on a vocal can help reinforce the beat while also helping fatten up the vocal due to the doubling-like effect of

the echo. (The simplest form of ADT -- i.e. artificial double tracking -- is simply a delay that is set short enough to appear to be someone singing almost, but not quite, in sync with the main voice.) I particularly like to configure both a short delay, such as a sixteenth note or eighth note, to help reinforce an undoubled vocal, and a longer delay, such as a quarter note, to yield an added sense of space.



The Sonitus:fx Delay included with SONAR Producer Edition is a stereo delay with a single tap per stereo channel. In addition to the standard feedback control, there is a cross feed control, which is like a feedback control, except it feeds the opposite stereo channel. Sonitus:fx Delay also eliminates most manual delay time calculations by synchronizing to SONAR's tempo and letting you input musical values, such as eighth and quarter notes. It worries about calculating the actual delay times, including recalculating them to follow any tempo changes in your song. All versions of SONAR also include the DSP-FX-based FxDelay and the traditional Cakewalk Delay. (There are mono and stereo versions of the latter, though neither is installed by default.) Cakewalk's Project5 also includes a tempo-synchronized delay called, perhaps obviously, Tempo Delay. For those wanting additional delay flavors, there are traditional offerings from a variety of companies, including dB Audioware, Waves, and PSP Audioware, among others.

Whereas delay is used for discrete echoes and rhythmic effects, reverb is used for providing more general ambience, be it natural, such as to simulate a room or other environment, or totally artificial. In a natural environment, a voice, or other audio signal, doesn't simply hit a surface and reflect back more or less as it was. Rather, it hits multiple surfaces at different times according to their position compared to the sound source. The materials used on those surfaces affect the reflections of the sound. Those initial reflections then bounce around further, as do their reflections, and so on. As time goes on, there are more and more signals bouncing around, hitting more and more positions on various surfaces, and these reflections begin to wash together into an overall ambience called a reverb tail. The reverb imparts a sense of space on the voice. A close miked vocal without reverb on it can often feel fairly unnatural, as we are used to hearing voices in rooms.



It seems ironic that we resort to all kinds of technology to simulate what occurs in nature. Nevertheless, a fair number of methods have been used historically to simulate reverberation, from springs to plates to simulation algorithms to analyzing how a room behaves when sent a known impulse and extrapolating the contributions of the room so they can be imparted on other sounds. As you might guess, this results in there being a wide variety of flavors of reverb available in plug-in format, sometimes even within a single reverb plug-in. SONAR 4 includes several plug-in reverbs in the box, including the Lexicon Pantheon reverb, FxReverb (based on the DSP-FX StudioVerb), and the traditional Cakewalk Reverb. SONAR 4 Producer Edition also includes the Sonitus:fx Reverb. If that doesn't provide enough variety for you, a sampling of third party plug-in reverbs includes Waves RVerb (Renaissance Reverb), PSP EasyVerb, Spin Audio RoomVerb, and the freeware SIR impulse response processor.

Here is a practical tip for reverb: We tend to like to hear vocals with reverb, but too much reverb can make a vocal unintelligible. Adding some pre-delay to the reverb, such that some of the direct vocal gets through before reverb signal kicks in, can help improve the clarity of the vocal. This can, in turn, let us get away with adding a bit more reverb since the pre-delay is preventing the reverb from getting in the way of vocal intelligibility.

Most of the processors and effects we've discussed, while capable of being used creatively, have been aimed at achieving natural vocal results. Of course, there is nothing saying you must aim for natural-sounding vocals. On the tame side, you might use EQ or a filter to make the voice sound like it is coming through a telephone. For some pop and electronic music there is the ever-popular Cher "Believe" effect, which is more or less pitch quantizing vocals, such as with extreme AutoTune settings. A heavy modern rock or industrial song might put a voice through guitar-like processing. A vocoder could be used to make an instrumental melody "talk" by imparting characteristics of the singer's voice on the sound of a synthesizer, or the voice could be used to drive a voice-controlled synthesizer. I know it's cliched to say this, but the only limit is your imagination, and perhaps your good taste. While it can be fun

to just play around with creative possibilities, it can also be easy to get carried away. One way to balance such tendencies is to look at creative processing from the perspective of what you are trying to achieve, and only then consider what processing is necessary to achieve that result.

Mixing Strategies and Issues

Okay, so we've got our comped/tuned/edited vocal, some thick vocal harmonies, and a general idea of how we want to process them. Somehow, though, when we mix the vocals in with the rest of the project, they just don't seem right. Either they're too buried in the instrumental tracks or they're too far out front, or maybe they sound like they're just plain disconnected, more or less in their own separate space. I imagine most everyone who has ever tried to mix a song with vocals has been in this position at one time or another. It is perhaps one of the most common challenges in mixing songs, perhaps because we humans do tend to focus so heavily on the vocals, and thus are more likely to notice something being even slightly off there than in other areas of the mix.

The bad news is there is no "one size fits all" solution. Each voice is different, as is each mix. Let me emphasize that last part: Each mix is different. The point being that what works with a given vocal in one context may not work for that same vocal in another context.

The good news is, if we recognize that the context will be different each time, even if the voice remains relatively constant, we can accommodate that in our mixing strategy. The key is keeping the vocal in mind at each stage of the mix.

One fairly common mix process is to start at the bottom and build up from there. For example, the kick drum might be used as the starting point, and that is made to sound solid before the next element, perhaps the bass, is added. Maybe the snare is added after that, then the rest of the drums. Then rhythm guitars to complete the basic tracks, and then piano, lead instruments, and pads. Finally things get to background vocals and vocals, but, by that time, there may be little space left in the mix. While there is nothing wrong with this bottom up approach, it tends to rely on the mix engineer's having enough experience to know what to leave out in order to leave space for what is yet to come. To put it another way, what sounds optimal at an early stage in the mix may not be leaving enough space for mix elements that will only be filled in later.

For a mix engineer who hasn't yet built up the experience as to what to leave out, bringing the vocal in earlier can help take the guesswork out of the process. Whether we go to the extreme of starting from the lead vocal, then add things from there, or we start with a few very basic rhythmic tracks, such as kick drum and bass, then add the lead vocal and build from there, the basic idea is to get the vocal in as early as possible so the mix can be built around it. Any time anything is introduced afterward, if the vocal all of a sudden starts sounding not quite so good, then it is the newest element which needs to be tweaked until the vocal is again sounding good. While this approach still leaves guesswork as to what other elements may be coming in later on, it gives priority to the vocal, almost guaranteeing a solid vocal because you aren't allowing anything else to detract from it. If you also stage the introduction of other elements to parallel their priority in the song, then odds are good you will improve the overall mix because you won't be committing to less important elements before you've gotten the highest priority ones nailed down. Kick and bass may still be important foundational elements, but other rhythm tracks may not be as important as the vocals and lead instruments, and, if that is the case, then why not give priority to the priority parts?

I've been primarily discussing this notion of priority at the philosophical and mix strategy levels, but let's get a bit more specific and talk about EQ. I mentioned earlier that I'm not a huge fan of EQ on lead vocals. The reason for that is, assuming a good match of mic, singer, and song, the voice without EQ on it will often sound ideal on its own. The reason we add EQ tends to be to help the voice sit better in the mix, but that represent a tradeoff if the voice's more natural state is our ideal. When we add other elements to the mix, though, some of those elements may step on the voice, and then the voice may need something to help it stand out in light of those new elements, right? An alternate strategy is to adjust those new elements so they don't step on the voice. That is, instead of EQing the voice to make it sit better once a new element has been added, EQ the new element to cut out whatever is stepping on the voice. In practice, I find this strategy is more likely to maintain the natural richness of the voice, without too adversely affecting the overall mix. Sure, some of the elements that need to be EQed so as not to step on the voice

may not sound optimal on their own. However, we are talking about an overall mix, not individual elements, and sometimes what is best for the overall mix is not optimal for individual elements. On the other hand, the vocal tends to be such an important part of the mix that something that is bad for the voice does tend to be bad for the overall mix.

A similar consideration can be applied to the spatial placement of the elements in the mix. Typically the lead vocal will be centered in the mix, as will be the kick, snare, and bass. Maybe lead instruments might also go there in exposed passages, such as an instrumental solo, but, by and large, we will likely want to keep other elements out of that space, letting the lead vocal shine through. While the snare may step on the vocal somewhat (we can always help that out with EQ), the kick and bass tend to inherently stay out of the way. Other instruments can be placed elsewhere so they also stay out of the way.

Back-to-front placement is often controlled via gain fader levels, but timbral characteristics also help give us the clues we need to get a mental image of where a sound source is. EQ might help by rolling some high end off instruments that are further back on the virtual soundstage. Or a harmonic exciter might be used to bring key elements, such as the lead vocal, forward without making them strident like adding EQ might do.

One particularly tricky aspect of vocal placement in the mix can be making the vocal and other tracks sound like they're in the same room. We often use different reverbs for vocals than for the rest of the tracks because we're going for a particular "vocal sound", which may or may not be natural, and the reverb that is required to get that sound may not be ideal for other tracks. Something that can help in this scenario is feeding a bit of the vocal into whatever reverb is being used with the instrumental tracks to help blend the vocal in with that "room". This may mean needing to turn the vocal reverb down slightly to compensate for any room reverb added, and the amount of room reverb added to the vocal would generally be slight -- just enough to give the impression it is in the same room as the other tracks, but not enough to distract from the vocal reverb's helping shape the overall vocal sound or to muddy up the voice.

Speaking of reverb, in general, an element of the soundstage that is closer to us has more direct sound, and less reverb, than an element that is farther away. Thus, if I want the background vocals to be placed further back in the soundstage than the lead vocal, I might give them more reverb. However, I also tend to treat the vocal reverb more as a part of the vocal sound than as an attempt at room simulation. Thus, I'd be more likely to give the background vocals more room reverb than the lead vocal, but the actual vocal reverb might be assigned more equally, or perhaps the lead vocal might even have more of it than the background vocals. It all depends on the context and what we are trying to achieve.

Wrapping Up

We've discussed many different areas of vocal production, though the list of areas we've covered is by no means comprehensive. There are numerous considerations that can, and often do, come into play when it comes to achieving excellent vocal recordings, and sometimes the sheer number of considerations can seem overwhelming. Hopefully you'll find at least a few ideas in this article that will be useful the next time you are wrestling with a problem vocal, or even an excellent vocal that just isn't sitting quite right in the mix.

What's New in AutoTune 4?



I've been a long-time user of [Antares](#) AutoTune, having started with version 1.3, then upgraded to version 3.0, which I've been using for awhile now. Both of those earlier versions seemed to do pretty much everything I needed to do. Thus, when AutoTune 4.0 was introduced, I was curious just what else might be coming, and whether or not that would make a difference in terms of either improving the quality of my results or speeding up the process of tuning vocals. Let's take a quick look.

Antares lists the following enhancements as being added in AutoTune 4:

Automatic Mode

- The ability to specify target note behaviors (Scale Note, Remove, Bypass) in individual octaves
- Virtual keyboard for target note control
- Improved Targeting option for problem vibratos
- Enhanced Vibrato functions for more realistic (or wacky) vibratos
- A Hold function to help analyze pitch errors
- Other assorted niceties (Dual Detune display, new Scale Edit buttons, etc.)

Graphical Mode

- Larger Pitch Edit Display
- Seriously expanded and reworked Graphical Editing Tools
- Multi-level Undo/Redo (up to 20 levels)
- Vibrato Scaling for modifying the depth of vibrato while maintaining its shape and character
- Streamlined navigation controls
- Cursor Pitch Display (back again by popular demand)
- Selectable Edit Display color schemes

Other

- 192kHz compatibility (host and hardware dependent)
- Selectable knob control
- Sleek new interface

Looking at the list, it would appear several of the enhancements target vibrato handling. Vibrato is always a tricky thing to handle when dealing with pitch control, because vibrato is, by its vary nature, a fluctuation in pitch. Thus, how do you determine what is intentional and what is simply pitch control problems? AutoTune 4 brings new features to the table for generating false vibrato and controlling real vibrato. The Improved Targeting feature (see also the main article) may or may not make things better for any given case, but it at least provides another option beyond the default handling in the event that handling didn't quite do the trick with a particular singer's vibrato. More options are generally a good thing when it comes to mind reading, which is essentially what automatic pitch detection and correction processors are attempting to do when they try to figure out what the singer meant to sing (versus what he actually sang).

The additional controls in the area of specifying target notes for automatic tuning purposes essentially help add more user control to this "mind reading" notion. Instead of just specifying notes in a single octave scale, it is now possible to specify different notes in different octaves. Let's say the particular vocal passage contains a middle C and a B above middle C, and the singer sings the B over a quarter step sharp. In AutoTune 3 and earlier, if you'd specified the notes C and B as being available in the scale, AutoTune would have tuned the very sharp B to a C. However, with AutoTune 4, you can specify that the C and B are available in the octave with middle C, but not in the octave above middle C. This would make AutoTune 4 tune that too sharp B down to a B.

These facilities are nice, but, having used AutoTune 4 on several projects now, I have to say my favorite enhancements are ones that are a bit harder to quantify. First, there is more space available in the pitch graph. I'm not sure how much additional space is available, but it feels like a significant improvement. Where that extra space comes in handy is in making it less likely that you'll have to move your cursor across screen boundaries while drawing pitch correction curves. If you do have to do that, though, AutoTune 3 users who were frustrated by how erratic the scrolling could be in the case of moving past the visible area of the screen could be will likely find AutoTune 4's behavior much more predictable, and much smoother. The net, at least for me, is improved productivity due to more time spent tuning and less time spent dealing with screen navigation.

I did find one idiosyncrasy with AutoTune 4. Specifically, I had several audio clips where AutoTune 4's graphic mode would not initially draw the pitch detection curve. The next user interface access after that would draw the curve, but would also get an odd error message saying, "Error: No error found." Hmm... That error message would continue to pop up with any further user interface actions, and it could sometimes be a challenge to regain control. This only happened on a small percentage of audio clips -- probably less than 5% -- and it isn't clear what the difference was between those clips and others where there was no problem. Antares did recreate the problem with my clips, but also could not figure out what was special about them. AutoTune 3 did not have the same problem with those clips. Also, if I slip edited the clips, then tried tuning them, that could make the problem go away, so there is a potential workaround in the fairly rare cases where this problem occurs. (Antares indicates I am one of only two users out of tens of thousands of AutoTune 4 users who have ever encountered this error.)

Overall, the improvements in AutoTune 4 are largely incremental. It's not likely that they will make the difference between your getting the results you want and not getting those results -- AutoTune 3 was highly capable in that department already. However, they may very well make a difference in how quickly you achieve your desired results. That is worth a lot when it comes to tedious process like manual tuning.

**Rick Paul is a songwriter living in Southern California. You can contact him at <http://www.RickPaul.info>.*