Review: Zero-G MIRIAM Virtual Vocalist
by Rick Paul - *12th October 2004 -*

Sampling and imitative synthesis technologies have come a long way over the past few decades. We've seen commercial products that can do a reasonably accurate job of recreating everything from orchestral instruments to rhythm guitars. Perhaps the Holy Grail of imitative synthesis and sampling, though, is the human voice. Sure, we've seen sample libraries with excellent choir "oohs" and "ahs", as well as libraries of canned phrases ranging from the musically useful to the absurd. Those really don't approach the flexibility and expressiveness needed for creating arbitrary vocal parts in a musical context, though. For example, it's unlikely I'd be able to put together a lead vocal for my latest song demo based on "oohs", "ahs", and canned phrases.

According to musical instrument heavyweight Yamaha, though, that is all about to change. Yamaha has introduced ambitious new singing synthesis software called Vocaloid, which they claim, "can generate authentic-sounding singing by simply inputting words and notes." While the software and technology behind the software belong to Yamaha, Yamaha is relying on commercial soundware developers to provide the voices through which Vocaloid sings and the products end-users will use to make their songs sing.

The first company to come to market with Vocaloid products for end-users is United Kingdom-based Zero-G Limited, one of the world's foremost providers of musical samples and sound effects. In fact, while the Vocaloid technology and software is still relatively new, Zero-G is already on its second generation of "Virtual Vocalist" products with MIRIAM, the product we will be reviewing here.

Basics

MIRIAM is a "Virtual Vocalist" (to use Zero-G's terminology), based on Yamaha's Vocaloid software and the voice of Miriam Stockley. Stockley is an established South African-born, London-based session singer and background vocalist who has worked with a list of notables including Tina Turner, Sarah Brightman, Queen, Elton John, and many more. She is also the lead voice in the successful new age Adiemus projects, has had her voice featured in songs in major motion pictures and television projects, and has released two solo albums as a singer/songwriter.

We'll get into the product architecture a bit more below, but, at the conceptual level, Yamaha's Vocaloid software (Version 1.0.5.12 was evaluated here) provides the engine, user interface, and certain generic vocal synthesis information, while the "voice database" embodies the singer-specific information. In fact, it is that voice database that differentiates the various Zero-G "virtual vocalists". Besides MIRIAM, LEON, a virtual male soul singer, and

LOLA, a virtual female soul singer, are also available, and share the same underlying Vocaloid software when multiple Zero-G Vocaloid products are installed on the same system.
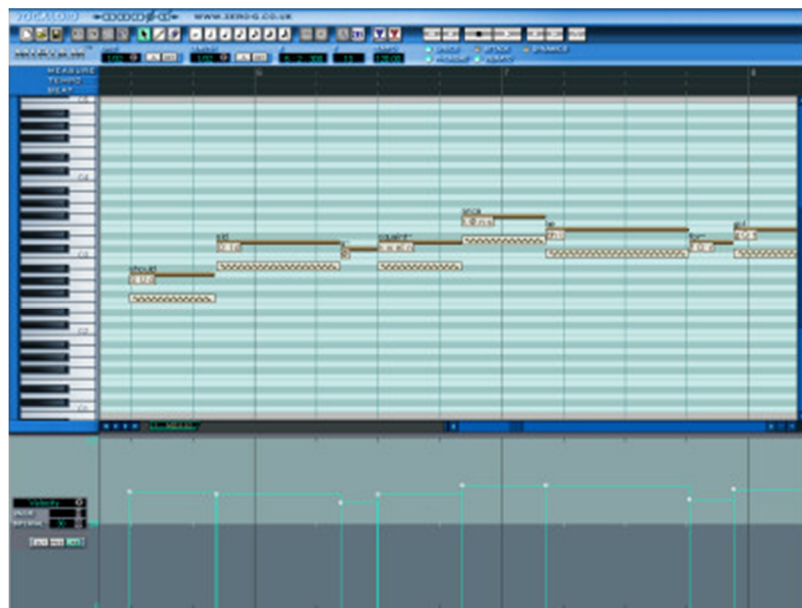
The MIRIAM box includes a 2 CD software set for Windows (Windows XP and Windows 2000 are supported) and a 100+ page owner's manual. The first CD provides the software, which includes a standalone Vocaloid Editor, a ReWire version of the editor, plus a play-only VST instrument with expression controls that can be adjusted at the DAW level. The second CD provides a roughly 9 minute video tutorial introduction to using the Vocaloid Editor, a number of example MIDI files demonstrating Vocaloid techniques, audio files demonstrating the results of using those MIDI files with MIRIAM, and audio demos of the other Zero-G Vocaloid products. The current suggested list price of MIRIAM is $279.95 (225 Euros or £149.95). The best street prices I found in a quick search around the net came in just under $250.

SONAR and Project5 users should note that, as of the present time, the VST instrument is of no practical use within these Cakewalk products. The MIDI sequences that drive the Vocaloid engine are completely based on NRPNs (Non-Registered Parameter Numbers -- a type of MIDI event), and the aforementioned Cakewalk products do not pass this information to VST plug-ins. This is true with both the current Cakewalk VST Adapter (V4.4.2 as of this writing) and the latest version of Tonewise's DirectiXer (V2.5). SONAR users will, however, be able to use Vocaloid's ReWire interface within SONAR. Unfortunately, Project5 does not support ReWire clients, so this limitation is more severe, and Project5 users will not be able to make practical use of Vocaloid at this time.

CakewalkNet is aimed primarily at users of Cakewalk products. However, due to the above-mentioned limitation of SONAR and Project5, I did some very basic testing of the VST instrument in Cubasis VST 4 (the version of Cubase that came with Garritan Personal Orchestra!) and with the Cubase SX 1.02 demo. Any references made below to the VST instrument were tested in Cubase. All other Vocaloid MIRIAM testing was carried out in SONAR 3.1.1 or in the standalone Vocaloid editor. Functional capabilities were also confirmed in SONAR 4.0, which arrived very late in the review cycle.

Installation, Software Protection, and Documentation

Installing the Vocaloid software and MIRIAM voice database is pretty straightforward. In case you have other Vocaloid software installed already, there are options to skip installing some of the components. One minor wrinkle I found was that, although the installation program allows customizing the location of the installed software, it temporarily copied the software onto my C: drive. Though I'd started the MIRIAM installation with 1.5 GB of free space on that drive, at one point I got a low disk space warning, signifying I had less than 200 MB of space left and prompting me to clean up. Luckily, the installation did not end up using all of that remaining space. I got by with just a nuisance warning, and the software could permanently reside on the drive I'd chosen for the installation, which had the requisite space to hold the permanent files (on the order of 1GB). However, if you don't have lots of space on your C: drive, this temporary space use during installation could prove to be an issue. (As a postfix on this, during the course of this review, I had a nasty registry corruption problem, not related to Vocaloid at all, which disabled Vocaloid and forced me to de-install and reinstall. I hadn't cleaned up my C: disk after restoring my operating system and getting things going again, and started the MIRIAM installation with about 1.1 GB of

available space on that drive. Even though I was installing MIRIAM to my F: drive, the installation ran out of space on C: and self-aborted. Cleaning up temporary files, compressing older files, and so on got me back to about 1.6 GB of space, and the MIRIAM installation then worked, leaving me just under 300 MB of headroom at the point where it was using the maximum space on the C: drive.)

Software protection is via on-line activation that is tied to the networking card on the system on which Vocaloid will be installed. Note that you do not have to have an Internet connection on the computer where Vocaloid will be used -- there is a way to get the activation information via another Internet-connected computer. However, you must have an Ethernet LAN card (or built in LAN device on the motherboard), and that device must be present in order to run Vocaloid. In this day and age of Internet pretty much everywhere, this method of software protection may not be a big concern for many uses. For example, most anyone with a DSL or cable connection, or, for that matter, most any relatively new consumer-type PCs that feature motherboards with just about everything but the kitchen sink built in, will likely have an Ethernet device of some kind. However, users with self-built PCs who specifically do not want a network connection capability on those PCs, or who only have modem access to the Internet may not like the idea of adding a network card simply to serve as a dongle. Also, there is the consideration of what happens if, for whatever reason, you need to change your networking card. Upgrading from one consumer PC with an integrated LAN card to another is one prime example of something that would cause a problem with the initial activation. Even if Yamaha or Zero-G will allow transferring the activation (the included license agreement did not specify, either way), there would at least be the hassle of having to reauthorize. (Zero-G indicates that Yamaha allows up to three activations per user. Thus, this would provide some built-in safety net to guard against LAN card failures or moving Vocaloid from one PC to another.)

The Vocaloid video tutorial is pretty basic, but covers the most essential elements in getting started with the Vocaloid Editor. I am somewhat disappointed that, when it covered some of the advanced controls used to further customize the sound of Vocaloid singers (more on these below), the values set appeared fairly arbitrary. It would have helped if there had been some explanation of how the values set related to an actual singing voice, assuming, of course, there were such relationships. However, it is possible that may be the nature of the beast in terms of the basic state of knowledge with respect to manipulating Vocaloid's controls to model a human voice. Also, that the tutorial covered those at all at least provides the user with a level of awareness that the controls are there and worth considering in order to get more expressive synthesized vocals.

The owner's manual is very well done and liberally laced with diagrams. A "Getting Started" section covers similar ground to the video tutorial, before a reference section takes over with detailed references for each operation within the program. There is also an appendix with phoneme (see below) charts, detailed MIDI implementation charts, and more. The one additional thing I wish it had is more application-oriented tips, especially as regards use of the expressive controls.

The good news is that there are additional Vocaloid-oriented resources available on the web, and the types of information available from these sources are expanding. Most significant among these is the Vocaloid User web site at www.vocaloid-user.net. The site includes additional tutorials, including some on the expressive controls mentioned above, updated phoneme charts, announcements of news relevant to Vocaloid users, and user areas including web forums and download areas for sharing Vocaloid MIDI files and audio files to demonstrate the results.

The Nature of the Beast

Before we start talking about the main components of Vocaloid and how to get started putting together a sung melody line, I just have to interject with a short note. This review is technically on MIRIAM, and Zero-G (and Miriam Stockley) no doubt dedicated a huge amount of time to developing the voice database that provides the raw materials for synthesizing MIRIAM's "voice", if you will. However, we will be spending a whole lot more time talking about Vocaloid in general than we do about MIRIAM in particular. The reason for this is that the user's interaction is primarily with Vocaloid. In fact, I might even maintain the user interacts exclusively with Vocaloid in most cases.

Now, you might be thinking that this isn't all that different from a sample library, where the only thing the user does with the library itself is load samples or multi-sample programs, and all interaction is with the sample player and maybe a MIDI keyboard and/or DAW? There are significant differences, though. For one thing, Vocaloid's recreation of a human voice is pure synthesis, although a high quality collection of vocal recordings was required for the frequency domain analysis required to generate the data used for synthesizing MIRIAM's voice. Perhaps more significantly, though, the results you achieve with MIRIAM, or any Vocaloid-based virtual singer available at present, will be influenced at least as much by your interactions with Vocaloid, and by its features and limitations. This will be more easily understood by the time you finish reading this review, but, in the interim, my apologies to Zero-G if it seems like we're spending an inordinate amount of time discussing Yamaha's product and not as much on what is specific to MIRIAM. It just so happens that we can't discuss MIRIAM without first covering Vocaloid in some depth.

Getting Up and Singing

There are two basic aspects of getting MIRIAM, or any Vocaloid-based virtual vocalist, to sing. The first is teaching MIRIAM the song -- i.e. the notes, rhythms, and lyrics. The other is fine tuning various aspects of MIRIAM's "singing technique" for a variety of purposes, mostly related to expressiveness. The first aspect and some elements of the second can only be accomplished in the Vocaloid Editor, which will be the user's main interface to creating MIRIAM's performance. Some elements of the expressive control are available through the VST instrument if you have a DAW, such as Cubase, which supports sending NRPN values to a VST instrument, but all of those elements can also be accessed from the Vocaloid Editor. Note that the ReWire interface to Vocaloid is, for most practical purposes, simply the equivalent of an audio patch cord between the Vocaloid Editor's output and an audio input to the ReWire host (e.g. SONAR), not a separate user interface. Short of using third party add-ons that provide their own interface for getting notes and words into the Vocaloid engine through the VST instrument (see sidebar on TazMan Audio Development's VocController), or receiving Vocaloid-ready MIDI sequences from someone else, users will start out in the Vocaloid editor.

The user's first task is to tell Vocaloid what notes to sing. The Vocaloid Editor provides a familiar interface for doing this -- a classic sequencer-style piano roll view. In case your love of piano roll note entry only slightly beats out the thrill you get sitting in a dentist's chair, I should note that you can also import a standard MIDI file into the Vocaloid editor. That MIDI file can be created however you prefer, be it playing it into a sequencer on a keyboard or guitar MIDI controller, entering it in a notation program, or whatever. You will likely have to do some cleanup inside the Vocaloid Editor, as the Vocaloid editor can only deal with monophonic parts on each of its 16 MIDI tracks (to be expected since vocals are monophonic). This is easily done, either by manually cleaning up note start times and durations or by selecting an entire melody, or just a range of offending notes, and using Vocaloid's Normalize Objects command, which truncates any notes that overlap the next note in the sequence. The Normalize Objects operation generally has the desired result since note start times are the most critical element in most phrasing.
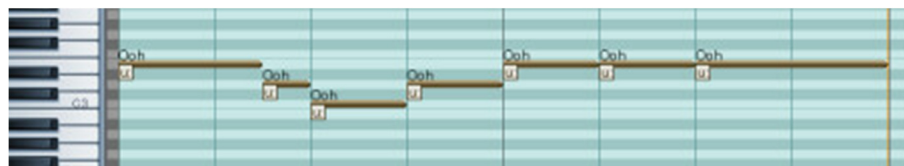
Let's take a look at this in action. In the small screen clip at left, I've entered the notes of the first line of the classic nursery tune "Mary Had a Little Lamb". At this point you can already get MIRIAM to sing. Just press the Play button and you're off. MIRIAM will sing the notes on an "ooh" syllable (click here to listen to the normalized MP3 clip), since you haven't yet "told her" what words to sing. What comes out is at least on par with most "ooh" samples you'd find on a self-contained keyboard, rack mount synth module, or sample CD.

The next logical step is to couple words with music, and it couldn't be easier -- simply type them in over the notes! If a word needs to extend over multiple notes, such as in the two-syllable word "Mary" or "little", use a hyphen to separate the syllables between notes. For this example that means typing "Ma-" over the first note and "ry" over the second.

It's worth mentioning that there are a few undocumented "tricks", in this area. Let's say you wanted to sing the word "love", obviously a one syllable word, but extend it over three notes. You'd enter "love-" on the first note to indicate the word will be extended, "-" (i.e. just a hyphen) on the second note to indicate you're still extending the current syllable, and finally "/" (i.e. a forward slash) on the last note to indicate the extended word is ending.

Another example is that there are some words we typically consider, and our dictionaries tend to agree, one-syllable words that Vocaloid expects to be entered as two-syllable words. In one project I was attempting with MIRIAM, entering the lyrics for one of my original songs, I came across two such cases. One was "fire". If you're like me, you consider this word to be one syllable, and sometimes maybe even pronounce it that way, but most of the time you pronounce it like a two-syllable word that might be spelled "fyer" (i.e. rhymes with "flyer"). Vocaloid expects you to enter "fire" as "fi-" then "re". Another similar word is "our". That one was tougher for me to come to grips with, because I often pronounce it pretty much the same as "are" (which Vocaloid does consider a one syllable word). However, Vocaloid considers it a two-syllable word, with the syllables being "o-" and "ur", and, another popular pronunciation of the word would rhyme with "power", which is decidedly a two-syllable word for most people.

Once you have the words entered, it is time to tell Vocaloid to try and figure out how to pronounce those words. Call up the Phoneme Transformation command off Vocaloid's Lyrics menu, and it will take its best stab. The screen clip at left shows what Vocaloid has guessed for "Mary Had a Little Lamb". The lyrics are on top of the notes, with corresponding phonemes below. Some of the symbols may look strange if you were expecting dictionary symbols, but hitting Play will, for the most part, produce a familiar sounding result. (Click here to hear it.)

I say "for the most part" because, depending on your accent, and Vocaloid's notion of word pronunciations, not to mention the pronunciation of the singer whose voice is being used underneath it all, things can sometimes sound a little odd. I might describe it as being akin to a non-native English speaker's speaking, or, in this case, singing, English. Also, Vocaloid has no knowledge of grammatical context and homographs (i.e. words that are spelled identically but have different meanings may be pronounced differently). For example, I might type, "there's a tear in my beer, " (the title line of a classic Hank Williams song) into Vocaloid, meaning I'd cried in my beer. Vocaloid doesn't analyze the context and cultural elements, though, and it comes back singing the word "tear" as if there were a rip in my beer.

The good news is, if you don't like what Vocaloid produces in the way of pronunciations based on your lyrics alone, you can change it. We'll briefly allude to more advanced topics in this area later, but, for the crying over a beer example above, there are two easy solutions. The first is to simply replace the word "tear" with "tier", which has the same pronunciation as the version of "tear" we have in mind. That may feel a bit odd if you've been conditioned to select the right spelling of homographs and homonyms in your writing over a long period, but, remember, your high school English teacher won't likely be looking at how you spell in the Vocaloid editor. It is about getting the right sound -- spelling be damned! (Gee, that felt good!) The second option is to leave the spelling as you had it (some things are just too ingrained), and directly changed the phonemes. Vocaloid transforms "tear" to "th e@" (I know, that looks really strange, but trust me, it would sound like "tare" as in, "rhymes with 'stare'"). I would probably modify that to "th I@" (pronounced like "tier"). Once you edit a note's phonemes by hand, Vocaloid will lock those phonemes so they aren't accidentally changed back to Vocaloid's default pronunciations the next time you run the Phoneme Transformation command. Tweaking the phonemes manually is also one of the easier ways to deal with differences in accent between regions. There are also some more advanced techniques that can be used to fine tune things once you've gotten things in the ballpark, especially on vowel sounds, at the phoneme level. Let's stay with the basics for the moment, though.

"I Hope You Appreciate How Hard I Work for You!"

Okay, thus far, we've got a melody and lyrics in there, and MIRIAM is already doing more than most vocal synthesis products to date have been able to do, and with less effort on our part. If all you need to do is demonstrate how words and music fit together for a co-writer, perhaps you're there. However, what you've got at this point will sound pretty stiff -- perhaps not quite "robotic", but not all that far from it, either.

Even at this level, though, it is worth taking a moment to appreciate what the work Vocaloid is doing behind the scenes. It has a dictionary that looks up phonetic pronunciations of words. That part isn't rocket science. What is more interesting is what Vocaloid has to go through to put those words together based on the data from analysis of Miriam Stockley's voice. We're talking about over 900 MB worth of data calculated by analyzing thousands of custom-produced recordings of Ms. Stockley's voice, by the way, so it wasn't only Vocaloid that's done some work on our behalf. Still, that work is over and done with now, and one can only hope that Ms. Stockley is back to singing real songs, instead of phoneme sounds, by now.

That 900 MB database of frequency domain data represents snippets of various phoneme transitions with different singing articulations (e.g. vibrato). Based on the words, notes, articulations, and more, which you have entered into the Vocaloid Editor, the Vocaloid software then retrieves the appropriate voice snippets, applies pitch conversion, then figures out how to splice together the right snippets to form the words you want, smoothly transition between snippets, add the proper inflections, and so on.

Yamaha calls the techniques it is using Frequency-Domain Singing Articulation Splicing and Shaping. Whatever you call it, though, if you start thinking about all the potential variables that come into play with respect to things like the words and phoneme combinations themselves, pitch possibilities, timbre-oriented possibilities, different ways of moving between two notes, ... Well, you get the idea. The notion of doing this in software, based on analyzing a professional singer's voice, including what she does in a number of singing situations, is pretty mind boggling. As we'll see later in this review, this notion comes into play in both the significant technical accomplishments Vocaloid represents and some of the issues that exist in the current state of this technology. It's worth stating up front, though, that what has been accomplished here is indeed impressive, independent of how that may or may not apply to given real life opportunities for using the current Vocaloid-based products.

"Once More with Feeling"

Once you've got the notes and lyrics entered, there are a number of ways you can go in terms of what you do next. The Vocaloid tutorial suggests adding articulations such as note attacks and vibrato. That is certainly one way you can go, and we'll get to that area shortly. There are also a number of expressive controls available that you conceivably could start using at any time, but, as we'll see below, those are best left for later in the game. My personal preference is to work with two basic areas somewhat in parallel, targeting changes that more directly relate to the basic musicality of the virtual vocalist's performance, leaving the more stylization-oriented elements for later.

The two areas I have in mind are, at the conceptual level anyway, phrasing and emphasis. In Vocaloid, phrasing translates roughly to note start times and durations. Emphasis, while also encompassing some areas we'll discuss later, mainly translates to note-on velocity. Note start times and lengths can be tweaked to emulate the way a real singer would phrase the line. Note velocities can be adjusted to emulate the way a singer uses volume for syllable emphasis and to emphasize the meaning of a phrase in the same way an actor might if speaking the same phrase. My experiments with Vocaloid to date have convinced me that these two areas are perhaps the most critical steps you can take toward making Vocaloid-based virtual singers sound more human, and more like a real singer (as opposed to someone with a monotone-type voice who just happens to have perfect pitch). I believe these two areas may be even more critical than adding articulations and experimenting with expressive controls. Those areas are also important, but doing them without getting the phrasing right, to my ears at least, yields a more robotic sounding result than working with both areas or even doing just the phrasing and emphasis changes without adding articulations.

The screen clip at right shows this technique in action with our favorite nursery tune. Note how some notes have been slid forward while others have been slid backward. I've also left a short space between "a" and the first syllable of "little" because some singers with less than great breath control might well pause there to catch a quick breath. Moving the syllables around caused me to need to lengthen some syllables. I did not need to shorten others, other than if I wanted them to leave a bit of space before the next note, because I used the Normalize Objects command mentioned above to take care of truncating any notes that had extended beyond the start of the next syllable.
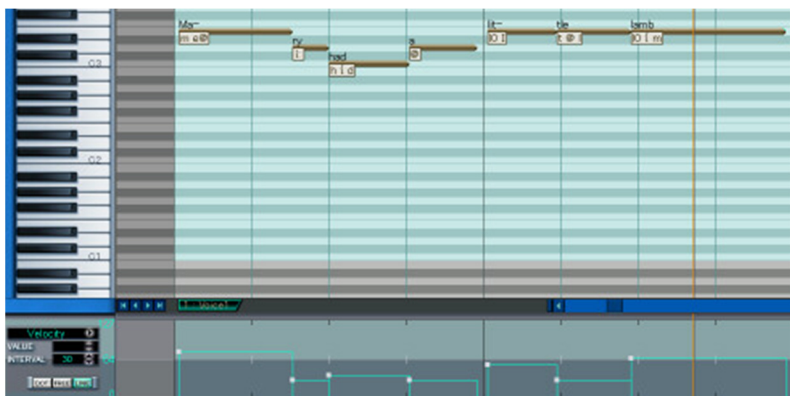
The bottom pane of the screen is where velocity values are modified. This is easily done by raising or lowering points associated with note on times. (Click here to hear the results.) One hint here is that if you're feeling mouse-challenged because of the small vertical space in which to move the velocity values in the default setup, you can change the location of the split between the two panes of the screen so the note pane takes up less space. This allows the lower pane, which is used for manipulating a variety of types of expression-oriented information (in this case velocity), to give you more room to maneuver in making your changes. Since you probably won't be working with a vocal range over about two octaves, there is plenty of space if you've got the Vocaloid Editor maximized on decent-sized monitor (I run a 19 inch CRT monitor at 1280x1024). I do wish Vocaloid would remember the pane size settings between sessions, though, as I find myself making the same change in most every session. (Besides giving more space to move around in the lower pane, having less of a range in the upper pane makes it easier to figure out which note is which at a quick glance.)

In dealing with the phrasing and emphasis here, there are two basic areas we are focusing on. One is how well the singer's rhythm fits the musical track. The other is how convincing the singer's phrasing and emphasis are. With respect to the musical side of things, I feel it is pretty important to hear the virtual vocalist against any musical tracks you may be using, unless you prefer to put together an *a cappella* "vocal" and add tracks later. (While I sometimes build tracks around a singer's performance, most of the time when I do that I've tracked the singer against either a piano track or some kind of generated accompaniment, be it from loops, an auto-accompaniment program, or whatever. Thus, there will typically be some degree of rhythmic references beyond just the singer's voice in the *a cappella* performance you get once the temp tracks are removed.) This means using the Vocaloid Editor in ReWire mode to allow synchronizing the virtual vocal with the tracks while still in the process of editing the tracks. While the Vocaloid Editor and SONAR work just fine together this way, there are a few limitations in using Vocaloid via ReWire that are, in my opinion anyway, among the biggest limitations in Vocaloid at the present time.

First, if you are using the Vocaloid Editor in a standalone fashion, it is easy to work on just a relatively small phrase at a time. This allows cutting the amount of synthesized vocal Vocaloid must render for you to hear your tweaks way down, so rendering is reasonably quick. Unfortunately, when you use Vocaloid via ReWire, Vocaloid starts rendering at the start of the sequence (including any musical introduction, pre-roll, etc.), thus having to do more work, and leading to commensurately long rendering times. Also, because of this, there is not really much point in breaking down a song into small phrases. This is because, as you move toward the end of the song, the amount of "pre-roll" that needed to get rendered would be just as bad as if you had lyrics there. Also, the extra overhead of needing to deal with multiple instances of Vocaloid configured through ReWire would probably not be worth the hassle just to save some time on phrases earlier in the song.

Second, the standalone version of the Vocaloid Editor has a capability called Play With Synthesis which is not available in the ReWire version. The Play With Synthesis capability will, if your computer is powerful enough and you have no more than 5 tracks in your Vocaloid MIDI file, let you hear your virtual vocal without having to render to audio first. That is, it does the synthesis in near real-time, so you can hear the sequence faster, check that your tweaks work, and either move on or try something else, as applicable.

The combination of these limitations mean you will have to choose between either being patient through potentially long rendering times or working in the standalone Vocaloid Editor. To give some sense of what I mean by "long rendering times" with the ReWire approach, a roughly 3.5 minute song I was working on with MIRIAM took on the order of 2-3 minutes to render on my Athlon XP 1600+ with 1GB of RAM. Needless to say, this kind of waiting can be quite frustrating if you need to use a lot of trial and error to get things "just right" (and my experience is that it does take a lot of trial and error to significantly improve results here). Using the standalone Vocaloid Editor approach means taking your chances with how things will match up with the musical feel of your tracks (or possibly doing some manual editing of the rendered audio track later on).

Thus, perhaps my biggest "wish list" item for a future Vocaloid update would be either making the Play With Synthesis capability available in ReWire mode or making dramatic optimizations in the way rendering is dealt with so that only what is changed needs to be rendered to hear playback. Being able to hear quick results would promote experimentation. At least with the current state of the Vocaloid products and technology, experimentation is critical for tailoring the output of Vocaloid to your tastes. This could be for pronunciation, trying to "get more feeling" in the performance, or even just to try and avoid some of the artifacts that Vocaloid's signal processing can have that detract from its believability as a virtual vocalist.

Icon Do More
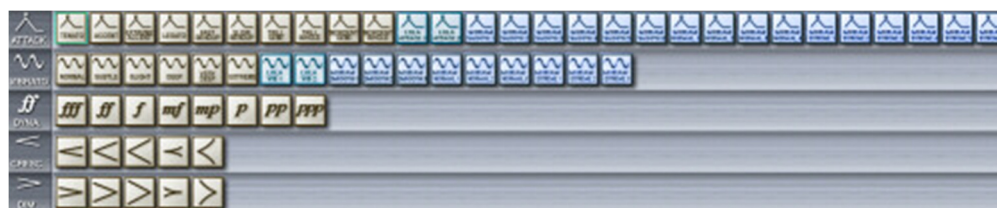
Vocaloid's Icon Palette (see screen clip at left) makes it easy to add preconfigured attacks, vibrato, dynamics, and dynamic changes to a virtual vocalist's performance. You can consider this library to be somewhat akin to a preset library of envelope shapes and parameters that are used to affect individual notes or the general vocal line that is playing. However, these shapes and settings are also modifiable for those who want more control, and this adds a considerable degree of expressive power.

To get started, simply drag the icon for a particular attack or vibrato setting to the note to which you'd like it applied. MIRIAM comes with a healthy assortment of attack and vibrato
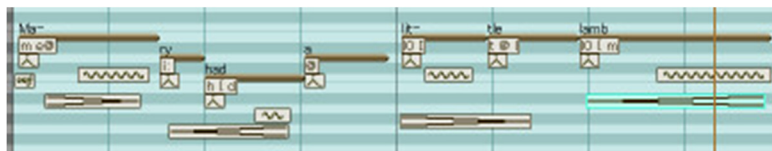
variations, ranging from the pretty subtle to the extreme. The key is picking a starting point that is in the general ballpark of what you want to do. You can modify it from there if need be. One simple modification for vibrato is changing where in a note the vibrato starts. By default, Vocaloid will start the vibrato at the halfway point of the note and keep it going until the end of the note. However, you can change where it stops or starts by simply dragging the starting or ending border of the vibrato icon to suit your tastes.

Even more control is available by double clicking on the expression icon. For attacks, you can change the duration of the attack and the depth. Vocaloid uses 64 as a default value for each parameter, meaning no change from the specified characteristic of the preset attack, with the overall range possible for each parameter being 0-127. Intuitively, lower values mean less of whatever, while higher values mean more of it. Vibrato offers even more customization potential. Position and duration control the starting point of the vibrato within the boundaries of the note and the length of the vibrato (which must keep it within the boundaries of the note -- don't worry, Vocaloid will keep you in line on this if you'd be inclined to mess up), respectively. These parameters correspond to the simple changes we described above that could also have been made with the mouse. More interesting are depth and rate curves, which allow altering, as you guessed, the depth and rate, respectively, of the vibrato over time. By default you start with a straight line at 64, reflecting whatever is in the preset. But you can change the value over time for dynamically changing vibrato. For instance, with depth, 0 means no vibrato, and 127 would be the maximum vibrato depth. With rate, 0 would be the slowest rate and 127 would be the fastest rate. For example, you could decide to

start off the vibrato slowly and with the "normal" depth. You might then speed it up and deepen it as the note goes on, but then slow it down and back off the depth before you hit the end of the note, in anticipation of the transition to the next note.

Dynamic markings work similarly, except they are not tied to individual notes. Be careful with fixed dynamic marks, which range from pianississimo (*ppp*) to fortississimo (*fff*), as they represent abrupt changes of level. In particular, unless you are going for some unusual effect, you will want to locate them either at the very start of a note, or, better yet in case you want to move notes around later, in the blank spaces between phrases. Each dynamic mark has a preset level and a range that corresponds to the general dynamic mark. If you double click on the dynamic mark, you can either change the level within the range for the mark you've chosen, or change to the range of another dynamic mark by typing a number outside the range. In the latter case, Vocaloid will switch to the mark within which that numerical value fits. You can also switch dynamic marks by choosing from a dropdown list, in which case Vocaloid will pick the numerical value in the center of the numerical range for that mark and you can modify it up or down from there.

Crescendos and diminuendos (a.k.a. decrescendos) are very useful for adding expression to a vocal line. You may remember that note velocity controls the emphasis of any given note, but that really only affects the start of the note. A real singer tends to vary the volume of a note over time, and also varies the volume of a phrase over time. Vocaloid starts you out with a number of crescendo and diminuendo icons corresponding to preset, line-based shapes. However, a real singer doesn't necessary just get louder or softer in a straight-line manner, and might get louder, then softer, then a little louder, and so on. Double clicking on a crescendo or decrescendo then going to the crescendo or decrescendo setup screen allows you to tweak the curve however you see fit. Just because you call something a decrescendo doesn't mean it has to get softer at all, no less get softer smoothly. For example, the screen clip to the left shows the diminuendo that lies under the first syllable of "Mary" in our continuing nursery tune example. I've left the starting and ending points as they were by default, but created a little swell near the beginning of the diminuendo to emphasize the vowel a bit. (Click here to have a listen.) Note that these dynamic changes are not tied to notes, but, because they happen smoothly, you don't have to be as concerned about where you place them. In particular, they have a range from -127 to 127, with a start value at 0, which means they start at whatever volume was in place where the crescendo or diminuendo mark starts. They also leave the volume at some value, be it the same level or something higher or lower. However, because Vocaloid does not allow dynamic marks to overlap, there is no chance of the kind of abrupt change you could get by having a static dynamic mark start in the middle of a note. The most you could do is use a very steep curve to ramp up or down quickly.

We're not done with expressive controls yet, but we have finished with the basic, but still very capable, controls available in the Icon Palette. Things will get more complex from here, but it's worth noting that you've already got a significant amount of expressive control without going beyond this point. It's also worth noting that even just tweaking away at this level to try and simulate the kinds of expressive things a real singer might do somewhat unconsciously can keep you busy for awhile. This begs the question of how far one should go with all this tweaking potential? Where do you reach the point of diminishing returns where you have to tweak an order of magnitude more to get only subtle improvements? I won't even attempt to answer these questions -- it would be presumptuous of me to suggest I have any answers, and those answers will likely depend on your intentions in using a virtual vocalist, as opposed to a real singer, in the first place. It is, however, useful to keep these sorts of questions in the back of your mind as we move forward with even more advanced tweaking possibilities then start considering applications for the Vocaloid technology in general and MIRIAM in particular.

A Pitch for Humanity

To this point, we've added a number of expressive elements to our brief Vocaloid track, but what we've got is still a bit too perfect. Even singers with great pitch control aren't perfectly on pitch all the time. While an off-key singer can make us cringe, a singer who was perfectly on pitch all the time just wouldn't sound human. What separates a singer with good pitch from a less talented singer is control over her use of pitch.

Vocaloid provides control over pitch bend and pitch bend sensitivity down in its Control Track area. Dramatic use of pitch bend can be used to simulate style-oriented techniques such as falloffs and bending up to a note. Less dramatic uses can create subtler effects, such as the kind of slight detuning you might get when doubling voices on a single vocal part or even just varying the pitch a tiny bit over time to humanize the synthesized vocal. Pitch bend sensitivity is variable in semi-tones up to plus or minus two octaves, though I think most typical uses, at least of the somewhat subtle flavor, will warrant a much smaller range to make it easier to create points on the pitch bend curve with precision. (Click here to hear the effect.)

Shaping the Voice

The Vocaloid parameters we've looked at to this point are only available through the Vocaloid editor in the standard Vocaloid-based packages. However, Vocaloid also provides a number of advanced parameters for shaping the voice, and these parameters can be manipulated in near real-time from the Vocaloid VST instrument in DAWs that can support it, in addition to being available for manipulation in the Vocaloid Editor. As noted above, SONAR's VST instrument support is inadequate for Vocaloid due to its not passing NRPN values through to VST instruments. Thus, in the interests of getting a more real-time feel for the effect these parameters have on the sound achieved by MIRIAM, I initially played around with these parameters in Cubasis VST 4. However, these controls are also available from the Vocaloid Editor's Control Track.

The Harmonics control does just what its name suggests -- adds or reduces the level of harmonics in the voice -- though it also has the side effect of making the voice louder or softer since harmonic content is being added or subtracted. This affects the sounds of vowels and voiced consonants (e.g. "b", "d", "n", "v", etc.). If this is as clear as mud, it is partly because, if you don't already know what harmonics do to a sound, what I've just said is probably

pretty meaningless. Even if you do, the effect is subtle enough to make it hard to explain. However, if you've ever worked with a harmonic exciter, or dealt with voices or instruments that have different levels of harmonic content, you probably have a sense of the effect increasing or decreasing a sound's harmonic content has. About the best I can explain it is that the voice seems to get a bit richer or more intense as harmonics are added, while it seems to get a bit purer and recede a bit as harmonics are taken away. Simple use of this parameter changes the nature of the voice in a relatively subtle way. More interesting use of the parameter is to vary it over time. For example, increasing the level of harmonics over the course of a held out vowel sound, then decreasing it slightly before the end of the note, provides a more dynamic sound than just keeping a static level. This better simulates what a talented singer might do such that her tone evolves over time. Another example might be that decreasing the level of harmonics over time could be used to simulate what happens when a singer starts running out of breath (i.e. with respect to vocal tone -- it wouldn't create the pitch variations that can also occur when that happens).

The Noise control adjusts the level of noise mixed with the voice. The Vocaloid documentation indicates that higher levels produce a harsher sound, while lower levels produce a purer tone. However, I'm not sure "harsh" is the right word here as it is somewhat ambiguous since there are other types of effects that could make a sound seem harsh besides what this control does. Also, the word "harsh" is overly negative, even for fairly extreme uses of the Noise control. It is more about what you are trying to accomplish than a subjective description such as this. To be more concrete, the effect you get by increasing the Noise control above the default value of 64 is somewhat akin to increased enunciation, and perhaps a bit of airiness in the voice. You can take this to an extreme, where it is almost like you're hearing the lip smacks and tongue noises of someone who is exaggerating the pronunciation of words. That might indeed be harsh sounding, especially if other Vocaloid parameters allow the noise element to be emphasized, such as if the Harmonics control is turned way down, in which case you get a result something like whispering, only with overly enunciated consonant sounds. However, mixed with elements that de-emphasize the noise component, it could be the just what is needed for better enunciation and clarity.

I will mention the Brightness and Clearness controls together because they are very similar, with their differences' being sufficiently subtle to make those differences hard to describe. The Vocaloid documentation indicates that the Brightness control affects the amount of high frequency content in the voice, with higher values creating a more present sound and lower values yielding a more mellow sound. As to the Clearness control, the documentation notes a similarity to the Brightness control, but indicates that the timbre of the voice is affected in a slightly different way, with higher values yielding a sharper, clearer sound while lower values give a more muted sound. If the distinction between those descriptions seems subtle enough, the distinction between their effects is at least as subtle, though there is an audible difference. As best I can distinguish between the two, low levels of brightness tend to emphasis the consonant content more, while somewhat veiling the vowel content. This can be an effect similar to a whisper, and the Vocaloid documentation does mention the potential of this control to simulate gently delivered or whispered phrases. The Clearness control, on the other hand, doesn't seem to affect the balance between vowel and consonant sounds to as great a degree. To my ears and way of thinking, it would actually seem to be more consistently clear throughout the range of Clearness control levels, whereas the Brightness control would seem to affect the clarity to a greater degree. Whatever the distinction, though, suffice it to say they are subtly different. Perhaps more importantly, you can use both controls together, be it having them both go in the same direction, having them go in opposite directions, or having some other relationship, even a random one, between them, to achieve different results. You can also use them statically to simply modify the tone of the voice, though using them dynamically will ultimately have more power for simulating the way a real voice behaves.

The Gender Factor control shifts the formant structure of the voice. Higher values shift toward a "more male" sound while lower values, which nominally shift things to sound "more female", tend toward a more munchkin-like sound. (Perhaps if starting with a male virtual vocalist instead of a female one, this would be less the case in that there would be further to go in that direction.) Even when shifting all the way toward the "male" direction, I did not get the impression MIRIAM sounded male, though, before getting to the relatively absurd-sounding degree, the sound did get into Mary Travers (of Peter, Paul & Mary fame) territory. By default, MIRIAM has what I might characterize as a light pop voice, perhaps somewhere in the territory of early Olivia Newton-John. In playing around with the various controls, I often found myself going for something at least slightly darker, even more toward Christine McVie territory, and the Gender Factor control, moved up slightly toward the "more male" direction, was quite useful for this purpose. To my ears, going in the "more female" direction was rarely flattering with MIRIAM, though a very slight move in that direction might be useful for achieving a characteristic more like a child's voice.

Next up are four sets of Resonance controls. Each set includes Frequency, Bandwidth, and Amplitude controls. As their name suggests, they set the frequency, bandwidth, and amplitude of the boost or cut in resonance for a particular band of resonance within the overall frequency spectrum. I suppose the conceptual notion might be something like simulating resonant areas within the body or mouth, and being able to tune those areas and make them more or less resonant to get different characteristics. I also have little doubt that, lurking somewhere within these controls, there is a key to capturing certain aspects of what a real singer might do when focusing her sound in different resonant areas of the body, such as the nose, throat, or head. The Vocaloid manual isn't very helpful in getting your arms around this area. The Vocaloid User Net web site has three tutorials on resonance that provide more information. For example, the first tutorial indicates the four sets of Resonance controls are actually modifiers for the first four, and most important in terms of their effect on the virtual "shape of the mouth", "formants" (which are themselves a sequence or resonances within the human voice). It goes on to say some key uses are to change the pronunciation of vowels and thicken up vowel and voiced consonant sounds, and it provides some example uses. For our purposes, I'll leave it to anyone who is interested in more on this subject to check out those tutorials. I will simply say that I've found that turning the amplitude of all four resonances down slightly -- e.g. from their default value of 64 to something in the 52-56 range, or even lower -- can help de-emphasize some of the synth-type artifacts present in the Vocaloid-based virtual singers. ([Click here](#) to listen.)

There's More

There are some in-depth areas of Vocaloid that I have intentionally not covered here. For example, besides tweaking phonemes to alter the pronunciation of a word, you can emphasize or de-emphasize consonant sounds in a word by drilling down to an even more detailed level of the phonetic pronunciation.. This level of control is very useful, but the more flexibility you add to what Vocaloid can do, the more work you are making for yourself if you actually want to use that level of control -- and, make no mistake about it, it is tempting. At what point, though, do you get to the level of work where it is just easier to go out and find a singer or to settle for "good enough"?

We'll talk a little more about real life applications below, but, before going on, it is worth at least brief mention of some additional facilities beyond what the MIRIAM package provides, which can augment or complement MIRIAM. TazMan Audio Development's VocController, in conjunction with the Vocaloid VST instrument, provides a level of real-time playability that is not present in either the standalone Vocaloid Editor or via the ReWire interface. Jasmine Music Technology's YV Enhancer (see sidebar), which is currently in prototype stage, will provide performance modeling technology to simplify the generation of dynamic performance control information which currently must be entered by hand. Then there are the earlier Zero-G Vocaloid Virtual Vocalists, LEON and LOLA, which can complement MIRIAM if you're looking for the voices of more than one virtual singer, for example, for a richer background vocal sound. Please see the applicable sidebars for brief overviews of these three areas.

"Where Do I Apply?

Before getting to my hands-on experiences with MIRIAM, it is worth spending a little time talking about potential applications of this technology. The following list may not be comprehensive, but some key potential uses that come to mind include:

- Songwriter work demos
- Songwriter final demos
- Master production background vocals
- Master production lead vocals
- Placeholders for productions-in-progress
- Sound sources for creative manipulation

Let's talk a little about what each of these notions might involve.

When I use the term "songwriter work demos", I am generally talking about a demo that is only meant for a few people to hear. Typically this would be the songwriter and any co-writers, who will use the demo while working on the song itself, and possibly as a reference when creating a demo for actual pitching of the song. Depending on the songwriter's relationship with his publisher, if there is one, or with other specific individuals he may wish to present the song to very early, specifically while it is still in non-polished form, the same demo might be used for very limited pitching. If so, it is probably mainly to get early reactions to the song before investing money and/or time in a more polished demo. Primarily, though, I am talking about a demo that will be used while the song is being written and rewritten, prior to the point it is determined that the song is worth pitching more widely. Thus, the demo is mainly for the benefit of a very few people who need early access to the song. One reason for using MIRIAM here might be that the songwriter truly can't sing. We're not just talking not having a voice you'd want people to hear on demos, as presentation isn't all that critical at this stage. Rather, we're talking one that can't carry a tune well enough to convey the tune, even to a collaborator or future demo singer who needs to learn the song to record it. Even a songwriter who could sing might consider using MIRIAM here, though. For example, perhaps it is quicker to set up MIRIAM and get something rough together than it is to set up a mic and record a real vocal, or maybe a male songwriter would rather hear the song with a female voice (or vice-versa). While working on rewriting, too, where the songwriter may use the computer, being able to just try changes out at the computer and actually hear them could be useful. The key for these types of uses is that vocal quality doesn't matter too terribly much in this case. What does matter is reasonable intelligibility of the lyrics and that it is relatively quick to input the melody, and that the melody itself will be accurate. The use here could include both lead and background vocals, but probably the large majority of uses will be for lead vocals only.

With "songwriter final demos", I am specifically talking about demos that will be pitched to others, with the idea of eventually trying to attract some other singer (or band) to record the song. This is not just a prototype version for a few close associates to check out while kinks are still being worked out. Rather it is an internal industry marketing tool that will be used to attract, in essence, investors in the song -- i.e. people who the songwriter hopes will spend their time and money to record the song and market that recording to the public in some way. It is not necessarily a recording intended to be heard by end listeners, but keep in mind that some industry types may have ears that are no more refined than end listeners, and they must be able to hear the quality of the song itself through any potential distractions. In this scenario, the lyrics will need to be clear and intelligible, especially if MIRIAM were to be used for a lead vocal. For background vocals, the key would be that the vocals convey any harmonies well enough to enhance, not detract from, the recording, and sound suitable enough for whatever the style of the demo may be. For lead vocals, in addition to the requirements for background vocals, it would be important to provide emotional impact that helps sell the song. Keep in mind, too, that some "songwriter final demos" will be very simple -- i.e. just a voice and a guitar or a voice and piano -- in which case a lot is riding on the vocal. In more fully produced demos, especially ones from a producer/songwriter who will eventually replace the demo vocals with those of the artist who is cutting the song, the tracks may be as important, but the song generally still needs a strong vocal to sell it. The style of the song will also influence the type of vocal that is acceptable.

When talking about master productions, I've broken the needs down into background vocals and lead vocals. Here we are talking about final product for end-listeners to hear, and this greatly intensifies the requirements for lead vocals, which are prominently featured in the track. While background vocals could also be relatively prominent, they may also be back much further, or even buried, in the track. Also, with background vocals, there is the possibility of augmenting any synthesized vocals with real vocals, while a lead vocal will generally only be from a single singer, so it will likely be all or nothing in terms of whether MIRIAM works for lead vocals in this environment.

When talking about background vocals in this application, the style will come strongly into play -- for example, the requirements for background vocals in bluegrass would be very different from background vocals for electronic pop. However, the key will be that the vocals blend in extremely well with the overall production, and do not have any noticeable artifacts in the final, mixed recording. This goes beyond the requirements for background vocals in songwriter demos. What you hear is what you get here, whereas, in songwriter demos, what you hear is only suggesting the potential for what you could get, and there is at least some assumption that those involved in the end recording can improve upon what is done in a demo.

For lead vocals in a master production, the considerations encompass all of those for the songwriter final demos and for background vocals in a master production. In addition, though, since we're talking about MIRIAM being the featured singer, not just a voice blended with other voices or buried in the background, the result needs to be fully believable for the style of the song. For most purposes (and we'll talk about some exceptions below), this means truly sounding like a human singer. This would set a very high standard, though different musical styles that typically allow for extremely heavy vocal processing (e.g. the all-too-popular pitch stair-stepping effect popularized by Cher's "Believe" among other mostly pop/dance recordings) may relax what synth-type artifacts may be allowable.

Whether a production is a master recording or a songwriter final demo, another possible fit for MIRIAM would be as a temporary, or placeholder, vocal to be replaced later by real vocals. In this case, the need is similar to that for a songwriter's work demo, where the intention is just to use the vocal while working on some other aspect of the song, or, in this case, production, or possibly as a reference for teaching real vocalists the song. The main value here is in the placeholder vocal's taking up the same sort of sonic space in the mix that would ultimately be occupied by the real vocals. This could make it easier for the producer/arranger to visualize (auralize?) those components of the end production while building other aspects of the arrangement.

One last potential application for MIRIAM that I'll mention is using MIRIAM as a raw sound source for creative applications, especially those where the original sound source becomes barely recognizable in the final mix. For example, people have long used talkboxes (remember Peter Frampton's "Do You Feel Like I Do" from the 1970s?) and vocoders to impart speech-like characteristics on instrumental sounds. Today, creative filtering plug-ins and software instruments, such as Antares kantos, which start with a voice to provide information elements that will be used in constructing the final sound, but which do not necessarily include the actual voice itself in the final sound, provide additional possibilities in this area. An example might be a synth sound that seems to talk, or to at least have some characteristics that give an impression of its talking. Another less extreme example might be a telephone voice, where the result sounds like a human on the other end of a telephone (or at least the caricature of telephones we typically get in a song). In such cases, because the end result will likely not be recognizable as a live human voice, whether or not there is an implication that it was at some point in the signal chain, the requirements for MIRIAM itself to sound believably human are much more relaxed. This is because the processing that will occur will ultimately dehumanize the sound anyway, and there is no real expectation that the end result will sound like a real human in the room with you. I might add that this particular application is one where real humans have historically been used, but not necessarily because there is a desire to have a real human involved in this manner. Rather it is because real humans have been able to do things that have been extremely difficult to synthesize to date, namely talk or sing. Elements of what they do fairly effortlessly are needed as raw ingredients in driving the sonic result, even if the singer's actual singing voice is not all that important. You could say this is an example where we've had no choice but to have a human doing the job of a machine (or piece of software in this case), rather than a case of an application where there is the potential to replace a human with a machine.

My own hands-on efforts with MIRIAM, and separately with LOLA (see sidebar), have attempted to cover much of this ground. As you might expect, some of those efforts were more successful than others. Let's dive in.

Time to Sing

Playing with new toys is always fun, but at some point it's time to get to the bottom line, which is what the product can do to meet your needs. I've intentionally avoided that aspect to this point because everyone's needs are slightly different, and, independent of those needs, understanding what goes into constructing a MIRIAM vocal with Vocaloid will be necessary to put the efforts, along with the results that can actually be achieved, in perspective.

While I typically try and outline my hands-on efforts in checking out a product in the order in which I undertook them, I am going to make an exception here. The main reason is that I believe it will make more sense to look at simpler, less demanding applications first, then build up to some of the more complicated, more demanding applications. You have to walk before you can run, even if that wasn't necessarily the order of my own efforts. A secondary reason is because, while this review is covering MIRIAM, I initially began my hands-on Vocaloid efforts with LOLA, with an earlier version of the Vocaloid software. Though I am not reviewing LOLA here, I did not get

to look at all the areas I will cover here directly with MIRIAM, and some of my earlier efforts with LOLA are relevant to the review. I will note any areas where the hands-on efforts were with LOLA, but it should be noted that I have verified that any behaviors reported here were checked as being consistent with how MIRIAM behaves. Thus, any issues of earlier generation Vocaloid-based products that were cured by the time MIRIAM came around will not be mentioned. (For the benefit of potential LOLA or LEON users, I might also note that updates to the Vocaloid software itself positively affected LOLA and LEON, too. Thus, the main differences that exist now between MIRIAM and the earlier generation of products are specific to the vocal databases and preset articulation libraries.)

Let's start with the songwriter work demo. During the course of my Vocaloid review efforts, I've written and/or rewritten several songs. While I sing my own demos, I have a little inertia problem for work demos. Basically, it is a hassle to get a mic set up and connected (mainly because my mixer is my mic preamp and I only use that when doing vocals -- the rest of the time it is covered in junk that I don't have room to store). If a song is finished and needs a rough demo, this isn't a huge impediment. If I know it will change due to rewriting, though, especially if I just need to demonstrate a small change in the context of a whole song, the thought of setting everything up to overdub a single line is a real momentum destroyer. Also, there is the issue that the line could be being overdubbed in a vocal that may be days, weeks, or even months apart from this session. That often means I'll consider whether to record the whole vocal over for consistency between parts, which would add to the overall time, especially when you factor in comping, editing, and such. Thus, what I've often done in the past is gotten lazy and just recorded an instrumental melody line, relying on my co-writer's imagination, and any clues I might give them in an e-mail, to figure out the fit of words and lyrics.

In this scenario, I've found the Vocaloid virtual vocalists to be a real blessing. While it will generally take me a bit longer to enter a Vocaloid part than just to record a vocal, or even several takes of vocals, it takes less time than the comping that goes on with even three takes (about the minimum I ever do). Also, in the event of rewrites (almost a certainty), any changes I make later can be made almost instantaneously with Vocaloid. This can be a big improvement in efficiency over the traditional route of setting up a mic, tracking the vocal, comping, and so on. Remember that we're just trying to get the song solid at this point, not worrying about presentation. Starting with LOLA, and more recently with MIRIAM, I have used Vocaloid for several early work demos of songs, namely to demonstrate fit of melody with words written by my co-writer. The result isn't a real human-sounding vocal because I do the minimum work possible -- i.e. just add notes and lyrics, with no attacks, vibrato, or other humanizing. Also, I generally just play the melody first in SONAR (did I mention I consider entering notes into a piano roll view to be similar to sitting in a dentist's chair?). I then create a MIDI file from that and move that into Vocaloid to add lyrics. This gets notes in quickly, but doesn't necessarily result in the optimal vocal phrasing. Of course, I could enhance things in Vocaloid if I wanted to, but the point here is just to get something together quickly so my co-writer can hear how the words and melody fit. Then we can rewrite both as needed, and using Vocaloid this way is a big step up over hearing just an instrumental melody. I can also incorporate changes very quickly during the rewriting process, and, at the end of it, when we're pretty sure the song is solid, I can record my own vocal with the final words and music.

The lines between songwriter final demos and master production demos blur somewhat for me, at least in the sense of the quality I strive for when doing a full tracks (i.e. as opposed to piano/vocal) demo. In particular, I am probably as demanding with respect to the lead vocal on a final demo as I am with a master production. In some ways, I am even more demanding in that I view being true to the melody, as opposed to having stylistic riffing in there, to be critical since the demo may need to suit multiple styles. Thus, if you don't mind, I'm going to talk about the lead vocal aspects of pitchable song demos alongside those for master productions and skip to background vocals for both purposes.

I love harmonies in general, and often create three part harmonies on my recordings. I generally double track those harmonies, but all the tracking and double tracking is with my own voice, and, while that sounds fine for some songs, it just isn't as rich as having, say, a trio of background vocalists tracking harmonies. In particular, the different timbres of different voices, coupled with slight pitch variations between takes and singers, helps sweeten the sound in a way that the same voice tracked over and over, independent of pitch variations, just doesn't do. Keep in mind, too, that time is always a consideration when talking about demos. Thus, especially for demo purposes, there is also the question of whether it might be good enough to just use virtual vocalists, be it MIRIAM or a combination of MIRIAM, LOLA, and/or LEON to speed up the production.

This is one of those efforts I'd mentioned above as having undertaken initially with LOLA. I'd already recorded a lead vocal and instrumental tracks for a pop/rock song, and was looking for some soulful background vocals, so I thought LOLA would be just the ticket. I initially played a harmony track into SONAR, exported a MIDI file, then read that into the Vocaloid Editor, which I synchronized to SONAR via ReWire. In Vocaloid, I cloned the track twice, and used the Vocaloid Editor's piano roll facility to delete the other two harmony parts from each of the three parts. (In hindsight, this might have been easier to do in SONAR's Staff View before importing the file, but it worked easily enough.) Next up was adding the lyrics to each of the parts (dead easy), setting up Vocaloid's mixer to pan each of the parts slightly differently, then hitting Play in SONAR to hear how the vocal track sounded against the rest of the song.

Okay, I skipped one important detail. In between hitting Play and hearing the track playing back was a wait of something like seven minutes for rendering the Vocaloid sequence to audio. The latest release of the Vocaloid software has improved rendering times somewhat, especially for the standalone version of the Vocaloid Editor. However, they are still quite long when using the ReWire connection. I have not revisited this particular song, but, based on my experiences with other songs, I expect the time should be cut down to 3-4 minutes.

The results? To be frank, I wasn't thrilled. I liked the texture the Vocaloid part added, but the pronunciations felt like I was listening to a foreign singer trying to sing English, and the overall feel was not terribly human. Note, though, on that last point, I didn't really expect it to be at this point because I hadn't added humanizing elements such as attacks, vibrato, and so on. The accent, though, was a bit more disturbing.

After experimenting more with humanizing elements and other expressive controls, and being frustrated by the long waiting times to audition the parts in ReWire mode, only to find the changes I'd made hadn't really worked out so well, I decided it was time to try another strategy. What if I forgot about trying to humanize or otherwise modify the Vocaloid tracks? Perhaps leaving them at the level of quick and dirty notes plus lyrics would work well enough if I added my own live background vocal tracks. I could balance the two any way that made sense. The only stipulation would be that the result had to be richer than I'd get with just my vocals while also avoiding obvious artifacts from the synthesized vocals.

Bingo! The Vocaloid tracks added nicely to the overall texture of my vocals, and I could balance the two together to minimize the artifacts. I'd also addressed some phrasing difference issues between the human and synthesized vocals via creative gating to tighten them against each other and the lead vocal. The style of production I was working on was such that there were already some electronic elements in there, and having a few less-than-human characteristics wasn't necessarily a negative at the overall mix level. If I'd needed to get more human with the Vocaloid results, I could have gone back and done more, and I've since learned a few process optimizing techniques that could help cut down on the rendering time concerns. In particular, after getting the initial parts done, I probably would have ditched SONAR for the moment and went to the standalone Vocaloid Editor. It provides a Play With Synthesis mode to get more immediate playback after making changes, and having the live vocal and instrumental tracks available to me while just doing humanizing edits wouldn't be all that critical.

Now, I mentioned I did this with LOLA, but I'm also very convinced that MIRIAM would work even better in this context. In particular, I find MIRIAM to be slightly more believable right out of the shoot, and the preset expression library (i.e. attacks and vibrato) designed for MIRIAM is also more comprehensive than that designed for LOLA. The tone would be different, and that might have worked better or worse on the particular song, but that's more a matter of specifics and taste than capabilities. Another option would be to combine MIRIAM and LOLA in this, be it having MIRIAM do some parts and LOLA do others or having MIRIAM and LOLA each do each of the parts (so, combined with my own vocals, each part would have been triple tracked). Some quick experiments I did on that front with MIRIAM and LOLA suggest that mixing the two virtual vocalists, especially using the triple tracking approach, could have been even more successful.

Based on my later experiments with MIRIAM and LOLA, I'm not as confident that using just one or both of these virtual vocalists for background vocals, especially in an exposed setting (e.g. bluegrass harmonies) would be satisfying. Extensive work with the humanizing controls, coupled with any external processing and mix tactics to hide or minimize artifacts, might work for some contexts. Perhaps, though, if we were talking about a song demo

only, where the idea is really to sell the song, not the recording, just using Vocaloid-based background vocals might be sufficient.

Before we get to more esoteric creative uses, what about the "Holy Grail" -- i.e. using MIRIAM for a lead vocal on a recording, be it a master or a polished, pitchable song demo? I currently have a production in the early stages of progress for a song that is really better suited to a female singer than a male singer, so I have begun trying to use MIRIAM to see if I can achieve a believable vocal performance for song demo purposes. If the recording truly came out good enough that I could use it as a master, for example, for film or TV soundtrack placement, I might consider that possibility. However, the real goal is to market the song to female singers who would record it and make it their own (and hopefully have a massive chart hit that sells gazillions of albums and all that good stuff). I might add, too, that there are some restrictions in the Zero-G Vocaloid license agreement that indicate uses of MIRIAM as the featured vocalist on a commercial recording or in certain film and TV contexts, may require additional licensing. I will assume for now, however, that, if the recording measured up, and if the opportunity were right, the business details could be worked out. In the meantime, there is the question as to whether MIRIAM can be believable enough for this use.

As I indicated, this project is really in its early stages, so I decided I would try and build the vocal early on, using temp tracks to give a rhythmic feel for reference, then replace the tracks later. Thus, I started out with some generated MIDI tracks playing back through a General MIDI-compatible softsynth module in SONAR. The next step was to use SONAR to play in the song's melody, export it to a MIDI file, and read it into Vocaloid, which was running via ReWire for synchronization with the tracks in SONAR, for adding lyrics. That first part went very smoothly, and, after a roughly 2-3 minute rendering wait between hitting the Play button in SONAR and hearing the MIRIAM track play back, I had an initial, but fairly mechanical sounding, vocal. So far, so good.

Next up, I did a little experimentation with note velocities, note-on timing, and note durations, and noticed that, even ahead of adding attacks and vibratos and such, this could make a fair amount of difference, maybe even an order of magnitude, in how natural MIRIAM sounded. Thus, my plan was to go through the whole song, fixing note timings and velocities to try and get as natural of raw phrasing as I could. I would also fix any little pronunciation quirks that I might be able to fix by either adjusting phonemes manually or by diving into the consonant balances within the phonemes for a syllable and tweaking those. I said "my plan" was to do this for the whole song, but, in reality, doing this was much slower going than I'd hoped. Part of this was the learning curve to find what worked and what didn't, and I got a bit more efficient on this as time went on. However, another part of it was the long rendering times in order to hear my results no matter how minor the tweak may have been. Working in the standalone Vocaloid Editor was not an option at this point since I wanted to hear how the vocal played against the musical tracks. I found it frustrating to have to either batch up a bunch of changes before listening to them or to have to repeatedly make a small change, wait a long while for it to render then play back, then adjust it further or move on to the next one. Keep in mind, too, that some types of changes, especially with expressive controls, can be interdependent. While I got a bit more intelligent about some of the changes as time went on and I could predict things better, there was still a huge amount of trial and error involved. This aspect of using the Vocaloid Editor via ReWire has to be my hands-down top beef with Vocaloid as it stands. In my opinion, Yamaha really needs to either get the Play With Synthesis capability working through ReWire or find some way to render only what is changed, not the full length of the song each time.

In any event, getting through the first verse and chorus of the song this way took me on the order of a week of elapsed time, probably spread over, on average, a couple of hours a day, five days a week. In case that sounds like an exaggeration, it actually may be, but, if anything, on the underestimation side. I just checked the file statistics from that song in SONAR, and it tells me I've spent over 38 hours working on the project. While there were no doubt times where the song was just sitting open idly waiting for me to finish with some interruption, the large majority of any real time spent on the song was working on the Vocaloid sequence. Just doing a bit of math, three minutes to render any change times even twenty different changes that needed to be tried before moving on would end up with an hour of time, not counting actual editing time.

By the time I had the first verse and chorus finished at this level, and sounding significantly better than the first iteration I might add, though still a long way from a human vocal from a professional singer, I'd decided to take two approaches to the next step. One was to send that iteration of the vocal for the song to Jasmine Music Technology to

run through their prototype YV Enhancer (see sidebar). I was very interested in what JMT's performance modeling technology could do toward automating the tedious work that lay ahead -- i.e. adding attacks vibrato, dynamics, and other control changes to get a more believable, more musical, performance out of MIRIAM. The second was to start down the path of making manual changes to try and do a similar thing -- i.e. make the performance more musical and more human.

I have to confess at this point that I never got anywhere on the second approach. A week passed between when I sent JMT the Vocaloid MIDI file and the point where they had a chance to run it through the prototype YV Enhancer. (Their running it through the enhancer, which is a partly automated, partly manually guided, process took less than 10 minutes. This could represent quite a few hours worth of manual work!) In the interim, I'd decided dealing with all the trial and error and long wait times was too frustrating. Thus, I started investigating methods for cutting down on the amount of rendering time. The bottom line with ReWire mode at the moment is there just isn't a good solution. Even if you break the song down into small phrases and only work with one at a time (decidedly non-optimal), Vocaloid will render any pre-roll in the song before the phrase starts. The farther you get back in the song, the longer the rendering will take. This may save some time on the front end, but eventually gets taking more and more, and the inconvenience of having to break things up to work this way becomes largely not worth it. What I did instead was work on a very short, separate example, specifically the first line of "Mary Had a Little Lamb" as used earlier in this review. This allowed me to spend quality time getting to know how various parameters affected the audible results while minimizing wait times by using the standalone Vocaloid Editor with the Play With Synthesis capability. More on that shortly, but let's get back to what happened with the first approach on the pop song demo's lead vocal.

JMT returned a MIDI sequence that had made significant musical and humanization-oriented enhancements to the very basic Vocaloid sequence I'd sent them. I would go so far as to say it was one or two orders of magnitude better than what I'd sent them, though it still wasn't entirely convincing as the vocal of a professional human singer. On the downside, their processing had changed all my manual velocity tweaking, and some of my timing tweaking. I decided I'd try and recreate much of that, and ended up opening a couple of instances of the Vocaloid Editor, tediously trying to figure out what I'd done in the one window and recreate it in the other window.

At this point, you may be wondering a few things: First, what if I hadn't spent all the time tweaking velocities and note start times and durations in the first place, and just sent the raw information for use with YV Enhancer? Second, was it really necessary after this to reinstate my changes?

On the first count, I'd also sent JMT the remainder of the song, which just had my raw, as in "not yet tweaked for phrasing and emphasis", lyric plus melody input along with the tweaked first verse and chorus. Listening to both outputs from YV Enhancer, the tweaked version was dramatically better than the raw version, which, despite the performance modeling changes, sounded much more robotic. So, yes, my manual tweaking up front was a necessary evil.

On the second count, my jury is out. I initially thought the YV Enhancer processing had tweaked the timing more than it did, but it turned out most of that was just note endings. I probably could have done a lot less editing than I ended up doing had I realized that because the other minor changes it made might not have mattered as much as I'd initially thought. Also, while I know the velocity changes I made were dramatically changed by YV Enhancer, and had made a dramatic improvement in the pre-YV Enhancer results, I'm not so sure trying to reinstate them helped. In particular, after having done the tweaks to get something resembling those changes back, my initial impression was the smoother flow of the YV Enhancer velocity levels may have been preferable to the more jagged flow of the levels I'd set. It made me wonder about the possibility that some middle ground might be the best option.

In any event, what I had at this point was a dramatically improved, thanks to the timesaving YV Enhancer-processed MIDI file, vocal track. Now it was time for me to look at some of the more advanced parameters that YV Enhancer wasn't doing anything with (at least in its present prototype incarnation) to see what further improvements I could make. While I could have gone back to SONAR and ReWire mode at this point, I knew I'd be doing a lot of trial and error tweaking, based on my "Mary Had a Little Lamb" exercise, so I decided I'd try two other tactics instead. One was to use the standalone Vocaloid Editor and its Play With Synthesis mode, and the second was to try using the VST plug-in in Cubase since most of the parameters I wanted to play with were tweakable in semi-real-time with the

VST instrument's interface. Eventually, I had to make any tweaks in the Vocaloid Editor, but my thinking was the VST instrument would make it much quicker to audition potential changes, and that thinking proved correct.

At this point the deadline I'd given myself for writing this review was approaching rapidly, so I decided I'd see what making some static changes to the remaining parameters could do. After much experimentation, I determined that turning down the amplitude for each of the resonance settings could help a fair amount in removing some of the synth filter-type artifacts that made MIRIAM sound somewhat mechanical by default. Playing more with the brightness and clarity parameters (YV Enhancer had automated changes in the noise and harmonics parameters) could further sculpt the vocal characteristics, and turning the gender factor slightly toward the male end also made desirable changes to the vocal character for this song. Overall, I would say those changes, coupled with a variety of vocal processing changes within SONAR (the vocal chain included a compressor, EQ, harmonic exciter, and de-esser, as well as sends to delay and reverb plug-ins), made another order of magnitude or so of difference in the result. I was starting to get to the point of feeling like that result might even be usable for demo purposes. However, I was at least somewhat concerned about the long hours I'd spent exposed to the MIRIAM part, and that the major improvements from the initial mechanical sounding part to where it stood now might have left my perspective a bit skewed.

A quick mix let me send a clip from the song to some colleagues for second opinions, and a break in the work while working on other parts of this review gave me a bit of space to come back to the mix with fresh ears. To hear the result they heard, check out this MP3 clip from "That Kiss" (words and music by Laban Johnson and Rick Paul, © 2004 Johnson/Paul, All Rights Reserved, Used by Permission). The verdict from both places was that my perspective had been slightly off, but the result was still a big improvement over previous samples. Some of the concerns were that the diction, mainly on consonants, was not clear enough and there were still telltale signs of a synthesized voice. The reactions also raised the question as to what would happen if I weren't shooting so much for a natural human singing voice, but something much more processed. We were thinking specifically of some of the stuff that has made it to the mainstream dance and pop arenas. I decided I'd try some further experimentation in that area before finishing off the review.

At this point, I thought it would be worth looking down two different avenues. One was just the heavily processed, but still human sounding voice. The other was the notion of using MIRIAM as sound and formant source for heavy duty processing, where the end goal is much more extreme or maybe not even recognizable as a human voice.

On the former count, my quick thought was to double up the MIRIAM part with the help of some delay, run both instances through AutoTune set to extreme settings (for an effect similar to the vocal processing on Cher's "Believe"), do a bit of different processing on each double (e.g. different compressors), then mix the two parts back together. The result was interesting, and probably a bit more believable because the explicit synthetic elements overshadowed some of the unintentional ones. A further experiment hooked PSP Nitro into one of the chains in the double, running through a large number of presets to see what kind of results might be achieved. Here things got even more interesting, showing some potential for the dance genre, as well as any electronica-oriented styles that use vocals, though it might be a challenge to balance the novelty of these effects with keeping someone's interest for a whole song. Another thing I tried, going back to just the single vocal, was running that through and EQ set up for telephone-style filtering. Possible pronunciation issues aside, I'm not sure I could tell the difference between MIRIAM and a real human voice with that heavy of filtering going on.

Moving further afield, into the realm of just using MIRIAM as an input sound source for really heavy duty sound processing or as an input into original synthesis, I stuck MIRIAM in front of Antares kantos. I ran through various kantos presets, then picked one as a starting point for manual tweaking. One of the big challenges in using kantos is keeping the singer's pitch stable enough to not get unintentional artifacts at pitch recognition time. While kantos has an AutoTune-like element in it to help with this, real singer's pitches aren't necessarily stable enough to have it always do the right thing. MIRIAM, on the other hand, has perfect pitch (unless, that is, you mess it up intentionally). As such, MIRIAM proved an extremely good input source for kantos, perhaps even better than a real human voice in some ways -- namely in the sense that it is more predictable.

Returning for the moment to the attempt to emulate a natural-sounding human voice, I alluded to my "Mary Had a Little Lamb" example earlier. You can judge the results for yourself via the audio examples accompanying this

review. However, my take on it is that there is probably more potential for simulating a human voice in Vocaloid, and with MIRIAM specifically, than can be used in practice with the current state of Vocaloid-based products. To tap that potential will take both more study into what really goes on in a human singing voice to know what needs to be done for certain situations, and expansion of the technology, be it through third party add-on tools or directly in Vocaloid itself. Also, for whatever elements remain to trial and error (hopefully at a higher level than what exists now), the ability to audition changes must be made instantaneous, or at least very nearly so. This is necessary to suit the kind of experimentation needed in areas where human judgement must be part of the process.

Closing Notes

It almost always tough for me figure out how best to summarize my overall reaction to a product. I strive to do this in a way that separates my personal needs out enough to take into account how the product might be useful to other people. However, I still want to reflect enough of my own experience in trying to use the product in real life, and to extrapolate how that might relate to what others in various situations might experience. This has never been tougher than with this review. On the one hand, I believe MIRIAM, and the Vocaloid technology in general, has some good applications right now. On the other hand, the way it has been marketed, and the way many people would likely want to use it based purely on their needs for a "virtual vocalist", are not truly in sync with the state of the product as it stands today.

Let's start with the bad news:

To be quite blunt, I don't think MIRIAM, or the Vocaloid technology in general, is ready for prime time when it comes to simulating natural human vocals that will be used in an exposed setting. Most of the examples I'd heard from other Vocaloid users, and on the Zero-G site itself, just did not feel convincing to me when MIRIAM, LOLA, or LEON were used as lead singers in anything resembling a natural voice setting. In trying to use MIRIAM (and LOLA) this way myself, I believe I learned a number of factors that can contribute to making Vocaloid singers more believable, or less so. I also believe I may actually have achieved a more believable result than I've heard elsewhere to this point, but it still wasn't convincing enough to my ears, no less to the ears of some of my less immersed-in-MIRIAM colleagues.

My reason for suggesting MIRIAM isn't ready for prime time as a lead vocalist replacement goes beyond just the end results, though, especially when talking specifically about use within the SONAR environment where the VST instrument is not an option. I want, and need, to hear MIRIAM in the context of the musical tracks. The time taken to render tracks for listening in ReWire mode is just to detrimental to the parallel need for lots of experimentation given the current state of Vocaloid and what it does and doesn't do semi-automagically. I do believe JMT's YV Enhancer could improve things significantly in this area once it is available. However, I also don't think that will be sufficient bridge the gap between where things are now and where they need to be to truly sound like a natural human vocal. If I'm correct in this, then it gets down to more need for experimentation, and this is where we need the faster auditioning enhancements in Vocaloid itself.

But let's move on to the good news, because there is plenty of it:

First, MIRIAM is worthwhile now when you start talking about less exposed uses, such as blending MIRIAM in to augment human background vocals. It may similarly make sense even just using MIRIAM in a background vocal situation where the vocals will either be buried sufficiently in the track or used in a genre where heavy processing fits the production. I am confident I will end up using both MIRIAM and LOLA when I work on productions where I want a full, rich background vocal sound beyond that which I can achieve by doubling my own voice. Sure, the slow rendering times hurt here, too, but the less exposed style of usage means many of the negative aspects of what you get with only a small amount of effort can be covered up in the mix, so there is less need for mandatory experimentation.

Second, when you get to heavily processed vocal sounds, be it the telephone filter type usage or an heavily processed electronica- or dance-type treatment, MIRIAM may well do the trick now, as long as you can get the

pronunciations and diction to the point where they suit your needs. This may be a "throw it against a wall and see if it sticks" effort. In particular, how successful this is will depend on the nature of the processing and how well it works with or against the unnatural aspects of what MIRIAM does with the level of tweaking you've seen fit to do. However, I do feel reasonably confident that it should find some application here. In fact, I fully intend to continue pursuing using MIRIAM for the lead vocal on the demo I mentioned above. The idea is that I'll first try using processing as needed to try and make MIRIAM work. The backup plan, though, is that, if I never get MIRIAM working to my tastes for the lead vocal, at least it's served as a good placeholder for building the tracks, and I can always replace "her" with a real singer if need be.

Which brings me to the third point. MIRIAM could already be valuable to songwriters and producers for the work-in-progress stages of their songwriting and productions, respectively. No one wants to have to hire a demo vocalist to come sing a song that isn't finished or to just be a temporary voice for a production that will later replace her voice with another singer. Sometimes it may make sense, especially in the production case, to do that anyway due to time considerations, but in the typical home or project studio case it will probably be undesirable for reasons of cost and scheduling. In this scenario, having a temporary stand-in "vocalist" could be a great help. And MIRIAM just might be a better singer than the songwriter or producer would be if that were the only other practical option.

Finally, I believe MIRIAM may make an ideal input source for vocoders and other processors that need sung words and/or human voice textures for input, but will output something that doesn't sound at all like a human voice, or which is so heavily manipulated that you can't tell if the original voice was a good singer, a bad one, or somewhere in between. After all, what singer really want her voice to come out sounding like a guitar or spacey synth? And some of these types of processors are finicky enough that having a very stable vocal source like MIRIAM may be an advantage over the more human characteristics of the real singer.

To sum up, I don't think MIRIAM, the Vocaloid-based virtual vocalist, is going to replace Miriam Stockley, or any real singer, anytime soon when it comes to a need for real human vocals from a pro singer. However, the leap it makes over past vocal synthesis products is pretty significant, and there is a lot of promise for the future. Some of that may even be applicable to current MIRIAM and the other current Vocaloid products since Vocaloid software updates can be made independent of the vocal databases. Even today, though, there are some fairly common uses for which MIRIAM is well suited, and I expect I'll be making good use of MIRIAM in my future songwriting and productions.

Playing Vocaloid VST from a Keyboard TazMan VocController

The standard Vocaloid VST instrument allows manipulating Vocaloid's expressive controls, such as resonance, harmonics, noise, and more in real time. It does not, however, let you play Vocaloid from a MIDI keyboard or standard MIDI sequence. Before you can do that, you'll need to create the actual melody and lyrics in the Vocaloid Editor. This is because the Vocaloid VST instrument doesn't read standard MIDI notes and controller values as inputs, but, rather, requires specially formatted NRPN messages. Enter the UK's TazMan Audio Development with their VocController VST instrument ($25 USD).

VocController provides the ability to play the Vocaloid VST instrument in semi-real time. "Semi-real time" because there is an inherent delay in Vocaloid, and Vocaloid also needs to know the length of notes in advance. VocController works with these limitations by providing a queuing system and selectable, musical delays such as a quarter of a measure or a full measure. Lyrics and their corresponding phonemes can either be entered directly into VocController or read in from a Vocaloid MIDI file. VocController will then couple those lyrics with the notes you play on the keyboard, going through the lyrics in order. Each lyric syllable can also have associated attack and vibrato presets attached to it.

In addition to tracking the MIDI notes you play and translating those to the NRPN values Vocaloid needs to see, VocController translates certain MIDI controller values, such as pitch bend, to allow for greater expression. It also allows automating volume, pitch bend, velocity, vibrato rate, and vibrato depth. Incoming MIDI note velocity can also be linked to Vocaloid's harmonics, noise, brightness or clearness parameters.

Unfortunately, SONAR users will not be able to take advantage of VocController's capabilities. This is because SONAR does not support sending NRPN values to VST instruments. In the interim, this capability is available to Cubase SX users, and may work with other VST platforms that both support sending NRPN values to VST instruments and provide the capability for one VST instrument to send MIDI output to the MIDI input of another VST instrument. I tried to make this work with the Cubase SX 1.02 demo, Cubasis VST 4, the Ableton Live 4.0.1 demo, SONAR 3.1.1 and 4.0, and Project5 1.5. Of these, only Cubase SX and Cubasis VST could get sound out of the Vocaloid VST instrument at all, and only Cubase SX provided the necessary MIDI routing from one VST instrument to another to allow VocController to drive Vocaloid.

From Notes and Lyrics to Performances YV Enhancer

Coupling words and notes is all well and good, but words and notes alone do not add up to a musical vocal performance. While the Vocaloid Editor allows manipulating a number of expressive controls, doing so is, at best, time consuming, and, at worst, an extremely tedious chore with much trial and error involved. The Vocaloid VST instrument provides real time access to some, but not all, Vocaloid parameters and you still have to understand how those parameters work to get predictable results.

Jasmine Music Technology (formerly NTONYX) has long been involved in providing performance modeling technology that breathes life into MIDI sequences by modeling aspects of real performances in various musical styles. Now their YV Enhancer, still in prototype stage, is aspiring to do the same for synthesized vocal performances created in Vocaloid.

The general idea behind YV Enhancer is that, within a given style of music, you can analyze what a singer is likely to do with various types of musical content. Once you know this, if you know the musical content, which, in this case, comes from a Vocaloid MIDI sequence, you can translate what a singer might do to Vocaloid's expressive controls. This will create a more musical performance. It will also relieve the Vocaloid user from the tedium of trying to do all the detailed control value entry by hand -- assuming the user even knows enough about what a singer would do in a given situation and how to translate it into Vocaloid control values. Some examples might include how velocity values might vary over the course of a musical phrase, how harmonics and noise might vary in the course of syllables of various lengths, and how pitch might vary as new notes are started and stopped.

According to JMT's Alexei Ustinov, YV Enhancer uses a "robot", called "PMStyle", which is "taught" by a style developer. During the "teaching" process, the style is tested on a number of musical examples with similar characteristics, until it can transform them automatically without major errors. Once a style is finalized, a user can apply the style to his own Vocaloid sequence, and can manage the style's work using a set of controls. The style itself is not intelligent in that it cannot "listen" to the result of the modeling -- it only reflects what the developer has "taught" it. But the user can manage the style's behavior according to taste, and, most importantly, can avoid the

huge manual editing chores associated with trying to create the performance control data by hand.

YV Enhancer is still in development at this point, so I couldn't try it out myself. However, JMT generously allowed me to send them a Vocaloid MIDI sequence where I'd entered notes and lyrics, and adjusted phrasing how I wanted it, but had not done any further enhancement. For example I had not added attacks, vibrato, pitch bend, or any of Vocaloid's other expressive controls. The "before" and "after" results were like night and day. YV Enhancer had enhanced the musicality and believability of the performance by at least an order or two of magnitude. This is definitely one development Vocaloid users will want to track!

Zero-G LEON and LOLA

Before there was MIRIAM, there were LEON and LOLA, Zero-G's Virtual Soul Vocalists. These were the first generation of commercial Vocaloid-based products, but, because they use the same underlying Vocaloid software as MIRIAM, they have also benefited from subsequent updates to Vocaloid. Both carry suggested list prices of $229.95 (USD).

While I have not tried LEON to date, this review cycle was originally begun with the intention of covering LOLA. However, the early Vocaloid software had a number of issues, and it was determined we would postpone the review pending an update that would correct a number of those issues. That update was longer in coming than initially expected, and MIRIAM was very close to ready by the time it arrived. Thus, the decision was made to shift the review to MIRIAM in order to reflect the state-of-the-art in Vocaloid-based products, including voice database lessons learned from experience with LEON and LOLA that could be applied in creating MIRIAM.

When I originally heard "virtual female soul vocalist" bandied about as a description for LOLA, I had visions of Aretha Franklin or Whitney Houston. However, the character of LOLA's voice is closer to Sade than either of those two. (Unlike with MIRIAM, the singers whose voices were analyzed to create the voice databases for LEON and LOLA chose to remain anonymous. Thus, we may never know who LOLA really is.)

For a listen to LOLA in action, check out this MP3 clip from "Undertow" (words and music by Mary Lou Sudkamp and Rick Paul, © 2004 Sudkamp/Paul, All Rights Reserved, Used by Permission). The clip blends the three vocal harmony parts covered by LOLA with my own live vocal doubles of those parts.

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