VORSIS Application Note

AN2008-06

Introduction to the VoiceMaster©
Distortion Management Algorithm

May 2008 – Jeff Keith

600 Industrial Drive New Bern, NC 28562 USA 252-638-7000
techsupport@wheatstone.com

Background

Today's competitive audio processors can do a remarkable job when tasked with processing stereo program material having a wide frequency spectrum such as that found complex music. The widely dispersed spectral content and generally wide stereo sound field in such material allows the various distortions produced by the processor to be at least partially masked by the audio itself.

When the same audio processors are tasked with processing more simplistic material such as live announcer voice, the rather sparse spectrum occupied by that type of signal does not allow the distortions generated by heavy processing to be masked as well. In fact, the post-processed quality of live announcer voice is a sensitive indicator of how well an audio processor performs when adjusted for competitive loudness.

Various schemes have been devised over the years to cope with the voice distortion issue. Some schemes might reduce the multiband limiter thresholds during live voice while others might decrease the drive level to the final clipper. While these might be overly simplistic descriptions of what might actually occur within processor products, the first thing that must be done before any 'voice' control signal is generated is to decide, and reliably so, *when* voice is actually present.

The approach that we took at Vorsis is multi-faceted. First, we have to decide whether or not voice or other mono programming is present at the processor's input. Then we have to decide what to do with that material once it has been detected.

Our approach does not modify the operating points of the multiband limiters or the drive level to the main clipper – during evaluation of other product's distortion control schemes we felt they were overly simplistic and sometimes sounded odd on mono signals compared to stereo program elements.

Our method detects and then processes voice (and other mono material) through a special processing sidechain that has been designed to sound good with that type of program content. The output of our scheme has a very well defined peak operating level (if desired) and can therefore *absolutely* control the *amount* of clipping on voice or other mono program energy simply because we can very accurately control the clipping duty cycle.

Is it Voice?

When deciding if voice is present at the input of an audio processor it is generally assumed that such an audio signal *must* be purely ‘mono’. Unfortunately it’s not quite that simple. Rarely during ‘mono’ programming are the left and right channels *exactly* in balance and also *exactly* in phase. What exactly *is* the difference then, between ‘voice’ and other programming such as oldies, that might *also* be in mono, and is it even important that we know the difference?

The answer is a surprising *yes* because voice and mono music such as oldies have completely different needs as far as their processing dynamics are concerned. To be successful the scheme needs to recognize different types of program sources and then adapt in a way that processes each in the most appropriate way *for that program type* and without negatively impacting the loudness or perceived quality of the program source.

Common Characteristics of Voice

In order to detect voice there are at least three things that we can look for within the incoming program material to determine if it might be present:

1. Is the waveform highly asymmetrical? Most voice energy has a lopsided energy distribution between the positive and negative going excursion of the audio waveform.

2. Is the signal primarily mono — are the left and right channels of nearly equal amplitude and also in phase? Unless intentionally altered by a special effect, voice is usually a *monophonic* signal.

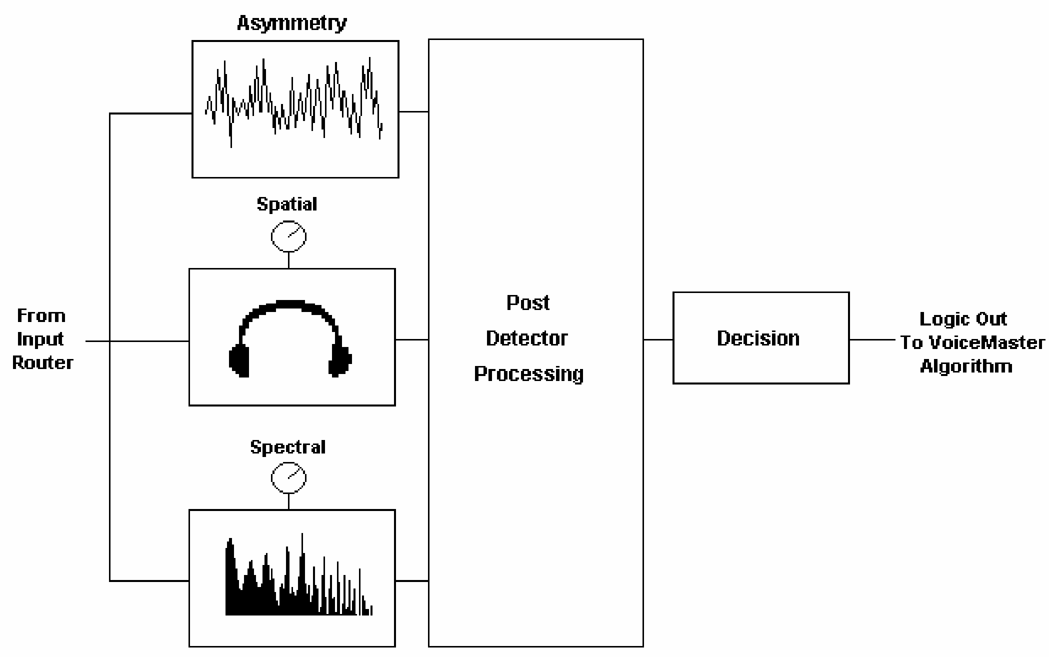
3. Is the audio spectrum fairly narrow? Again, if not altered by a special effect, voice has a fairly narrow frequency range – narrower in fact than most other types of common program material.

We might also use the cadence of the program as another identifier for voice however our research has shown that this is not as reliable as other methods.

By defining the above characteristics at least in some general terms we've narrowed down three likely identifiers for voice:

- Waveform Asymmetry
- Spatial Characteristics
- Spectral Characteristics

Now that some primary attributes have been defined, how do we find them in the input audio? One way is to use special detectors that have been tuned to look for each characteristic and then gang them by digital logic in a way that creates a reliable voice detector. See the simplistic diagram below for how the three main detectors in our VoiceMaster algorithm are arranged. A more comprehensive diagram can be found later in this document.



VoiceMaster - Basic Logic Diagram

We define the output of an unsatisfied detector as a logic "0" with the output being a logic "1" when it *is* satisfied. It is important to recognize that different types of program material *may* satisfy one or more of the detectors at any instant, and therefore it is critical that we ignore the satisfaction of *individual* detectors (or pairs of detectors) at *all* times.

Note that in the above diagram there is no user control shown for the Asymmetry detector. That detector has been designed to be self tuning and therefore requires no user adjustment in order to operate correctly.

The following table illustrates the role of the three detectors in assisting the algorithm in deciding whether or not voice energy is present. There are eight possible logic conditions (0 through 7):

	0	1	2	3	4	5	6	7
Asymmetry?	0	0	0	0	1	1	1	1
Spatial?	0	0	1	1	0	0	1	1
Spectral?	0	1	0	1	0	1	0	1
Voice?	No	No	No	No	No	No	No	Yes

Only when *all three* detectors are satisfied will a logic signal instruct the VoiceMaster algorithm to alter the processor's behavior to the incoming program material. Whenever this logic signal is a logic "1" the output signal of the 31-band algorithm will be processed by its own user-adjustable look-ahead limiter and four-band full parametric equalizer.

What about Mono Music?

While the VoiceMaster algorithm is quite reliable at detecting and processing voice, it is also designed to detect and *adeptly* process mono material such as oldies *music*. No other broadcast audio processor that we are aware of today has this capability.

The primary reason why a different style of processing might be desirable for mono music is that stereo programming is more tolerable of 'clipping' artifacts than mono material because the stereo sound field in normal stereo programming usually helps to acoustically mask certain processing artifacts. Mono material unfortunately does not benefit from 'spatial masking' and therefore can be more temperamental to process 'competitively'.

The VoiceMaster algorithm has been designed to be extremely adaptable to all kinds of monophonic program material and its time constants and spectral and spatial detectors were made fully adjustable in order to enable tuning for and operation under the widest possible variety of competitive programming conditions.

In summary, VoiceMaster processing is not only quite adept at processing live voice, users will also find that it is also quite suitable (and probably preferred) for processing mono programming.

VoiceMaster Control Overview

Remember that the logical location for the VoiceMaster algorithm is *after* the 31-band FM limiter and *before* the main clipper. This location for VoiceMaster happens to be the best one for maximizing loudness while minimizing clipping on the type of program material that least tolerates it.

Remember also that VoiceMaster is not a simple "…gain control in front of the clipper…" as might be found in other products, but is a complete processing chain specially tuned for mono program material.

The following graphic shows the Input Menu of the AP2000/FM2000 processor along with the control area for the VoiceMaster algorithm.

The operating controls for VoiceMaster are as follows

AutoMono – arguably the most important control because when AutoMono is Off, VoiceMaster is also off! AutoMono is a required part of the VoiceMaster algorithm and it needs to be enabled because several of its detectors are utilized for certain parts of the decision making process.

Controls within the VoiceMaster specific area:

Enable – this is the second most important control besides AutoMono because when the VoiceMaster Enable check box is not checked VoiceMaster is Off.

Spatial – this control adjusts the sensitivity of the algorithm that monitors the difference in energy between the sum (L+R) and difference (L-R).

Not simply a level detector, it's a spatial detector that measures the ratios of peak *and* average energy of the L+R *and* L-R in selected portions of the audio spectrum. It then outputs a logical decision when the output of the Spatial detector has reached a user-set amplitude when compared to energy occurring in the L+R.

Spectral – this control sets the comparison level of the ratio *differences* between certain frequency-domain signals within the L+R and L-R. The Spectral control is used to set the sensitivity of the algorithm that compares the spectral energy inside and outside of the vocal frequency range.

Lim. Drive – this control sets the Drive level to the VoiceMaster peak limiter which is responsible for processing voice and monophonic material when the VoiceMaster algorithm is active.

Pk. Ceiling (Peak Ceiling) – this control sets the absolute peak output level of the VoiceMaster peak limiter. This control is important for making the loudness/distortion tradeoff on voice and other monophonic material.

The Pk. Ceiling control is calibrated so that when it is set to "0.0dB" the peak output level of the VoiceMaster limiter exactly matches the clip threshold of the main clipper when the VoiceMaster limiter control is also set to its fastest attack time (0.5mS). What this means is that when the VoiceMaster limiter control is set to its fastest attack time (0.5mS) and the Pk. Ceiling control is set to "0.00" it is virtually impossible for voice or other mono material to exceed the main clipper threshold

Lim Atk – (Limiter Attack) sets the attack time of the VoiceMaster limiter. Faster attack times (lower numbers) increase the obviousness of processing and reduce the number of audio peaks escaping the limiter. As longer attack times are used, the Pk. Ceiling control will need to be reduced to accommodate the additional peak energy escaping the algorithm that reaches the main clipper.

As explained above, the Pk. Ceiling and Lim Atk. settings control <u>how much</u> voice or mono energy is permitted to reach the main clipper – when set to “0.00dB” and “0.5mS” respectively voice and mono material may *reach* the main clipper threshold but will never *exceed* it. This results in extremely clean and clipping-artifact-free processing of that type of program material.

Lim Rls – (Limiter Release) sets the release time of the faster limiter time constants within the VoiceMaster algorithm. Faster release times (lower numbers) increase the obviousness of processing without affecting the number of audio peaks escaping the limiter while also increasing the loudness and density of mono material. Settings above approximately 150-200 milliseconds may cause audible ducking when the limiter is in operation, therefore it is recommended to use settings below approximately 150 milliseconds.

AGC Atk – (AGC Attack) sets the attack time of the slower time constant that rides along with the VoiceMaster limiter control signal. The purpose of this control signal is to reduce the obviousness of processing by creating a longer-term average gain platform from which the VoiceMaster limiter action is based.

As the AGC Atk control is set to faster settings, more AGC action will occur which will lessen the possibility for the audible ducking warned about in the preceding paragraph concerning Lim Rls (Limiter Release).

AGC Rls – (AGC Release) sets the release time of the slower time constant’s average gain platform. Higher settings slow the release action down which reduces the ‘obviousness’ of processing. The higher the setting of the AGC Rls control the more gentle the processing sounds, even though it is still aggressively controlling the peak energy in voice and mono program material.

It is recommended to use the “Limiter” attack and release and the Peak Ceiling controls to set the amount of clipping permitted on mono material and then use the AGC Attack and Release controls to set the desired processing obviousness. There is enough range in the VoiceMaster operating controls to allow mono material to sound any way the user wishes from ‘…a little extra processed…’ to just like the stereo program.

And… Voice Asymmetry?

As mentioned previously we have not provided a control for “Asymmetry”. This is because the algorithm is completely self-tuning and does not require user adjustment in order to compensate for different characteristics of the incoming program material. The algorithm is able to appropriately adjust itself for waveform crest factors between 3dB and 20dB.

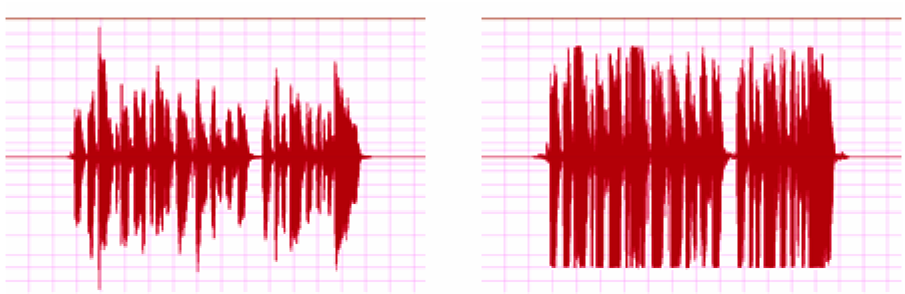
VoiceMaster Operation

The following table lists the suggested starting points for the VoiceMaster operating controls. They have been set for good tradeoffs between voice/music sensitivity, processing obviousness on voice and mono material, and best behavior while transitioning between processing stereo and monophonic program material.

AutoMono	IN	*Pk. Ceiling*	-1.5dB
AutoMono Threshold	-27.5dB	*Lim Atk*	16.0mS
Spatial	-27.5dB	*Lim Rls*	78.0mS
Spectral	-3.0dB	*AGC Atk*	130.0mS
Lim Drive	-0.5dB	*AGC Rls*	880.0mS

With the above settings both voice and mono program material such as music will be carefully and unobtrusively limited by the VoiceMaster algorithm with only a hint of clipping distortion permitted by the main clipper.

The graphic below compares how a voice waveform might look when it is unclipped versus being clipped. Note that although clipped audio sounds louder, the distortion created by the clipping process adds harmonic energy that will give it an edgy sound. This may or may not be desired, depending on the station, its format, and its competitive needs.



Unclipped Voice Clipped Voice

If *more* clipping is desired in order to add a certain *bite* to the sound of voice and mono material then the Pk Ceiling control may be adjusted to a slightly higher setting – perhaps to -0.5dB or even 0.0dB.

If it is desired to have *no* hint of clipping on voice and mono material, then the Pk. Ceiling control may be lowered from our recommended settings to perhaps -2.0dB or even -2.5dB depending on how aggressively the main clipper drive control in the FM Limiter menu has been set.

The final settings of the VoiceMaster controls, especially those related to the Attack time and Peak Ceiling values are a matter of personal taste and the competitive requirements of the station. Your ears are *always* the best tools to use for making final judgments!

Note that the Lim Atk (Limiter Attack) control has an important role in how the Pk Ceiling (Peak Ceiling) control may be set for certain program material. The lower the setting of the Lim Atk control (the faster the attack time) the fewer the audio peaks that will escape the VoiceMaster limiter. This means that the Pk Ceiling control may then be set higher <u>if</u> the Lim Atk is <u>also</u> set for faster attack times.

It is worth repeating that the *AGC portions* of the VoiceMaster algorithm have been provided in order to enable the end user to tailor the overall sound of mono processing compared to that which occurs with stereo program material.

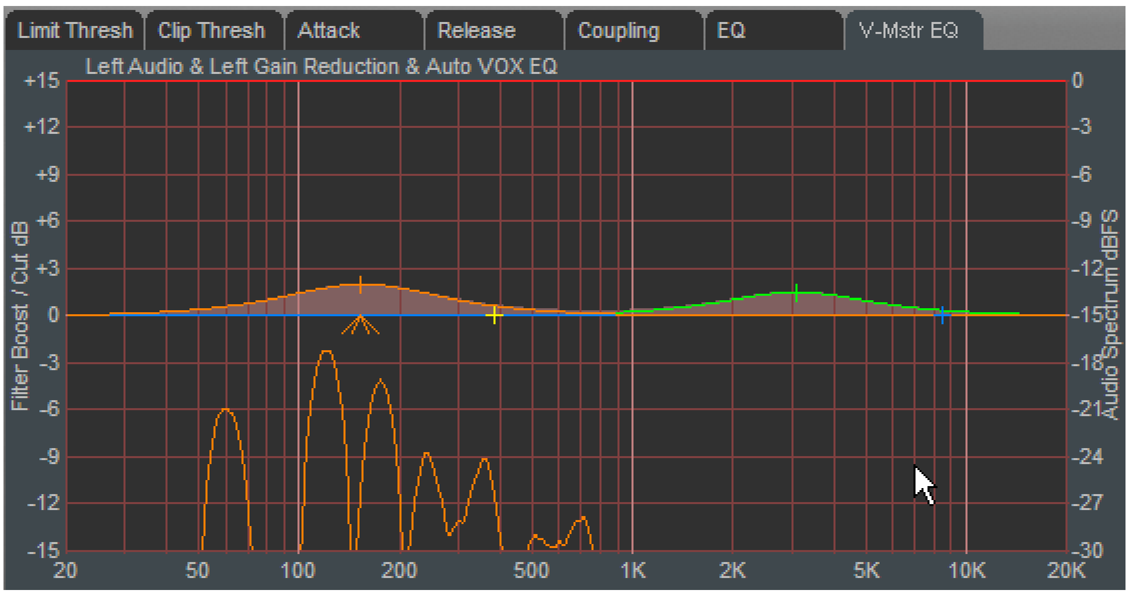
With our recommended settings the difference between the sound of stereo and mono program material is virtually indistinguishable as far as overall on-air loudness and program texture are concerned. Also, the perceived distortion on mono material is subjectively identical to that which might be perceived with most stereo programming. This lends a definite ‘big station’ consistency to any station’s on-air sound.

The above statement holds true unless radical equalization differences exist between the stereo and VoiceMaster signal paths. Toward that end we recommend that equalization on the VoiceMaster path be used sparingly and that boost and cuts be kept on the order of 3dB or less.

Please see the next section on VoiceMaster equalization for shaping the sound of the VoiceMaster signal path to your requirements.

About VoiceMaster Equalization

Within the VoiceMaster limiting algorithm is a four-band, full parametric equalizer whose sole purpose is to allow *subtle* shaping of the overall sound of the mono path compared to that used for processing stereo. We carefully chose the word *subtle* because any equalization done at this point in the signal path should be for minor corrective purposes only and therefore some care and common sense is warranted.



VoiceMaster Parametric Equalizer Control Screen

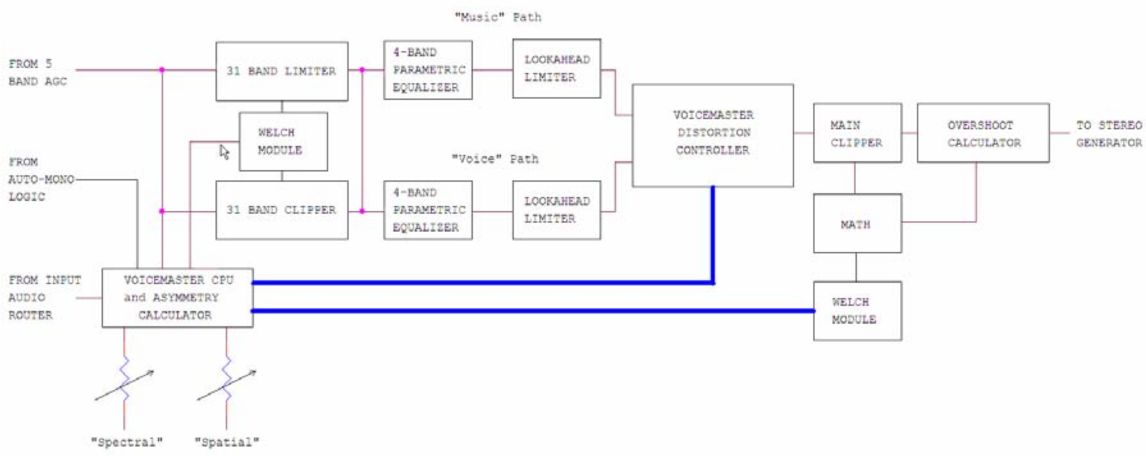
Our general guidance on VoiceMaster ‘post-equalization’ is this:

If more then 3dB of VoiceMaster boost or cut equalization is required to ‘balance’ the sound one should look carefully at how the previous sections of the processor, namely the multiband AGC and parametric equalizer, have been tuned. If the settings in these sections appear to ‘make sense’ and ‘stereo’ material sounds just fine, then it may be prudent to extend the investigation to the signal chain that is prior to the processor.

Note also that we recommend the use of fairly low Q settings (wide bandwidth) so as not to cause the sound of the VoiceMaster algorithm to be too ‘peaky’.

VoiceMaster Internals

The following diagram reveals *some* inner workings of our VoiceMaster algorithm (minus certain omissions of a proprietary nature).



We shall not delve deeply into the design of our algorithm; however studying the above diagram will lend insight into how the algorithm seamlessly processes both stereo and monophonic program material while maintaining a subjectively balanced distortion profile on the air.

Our VoiceMaster algorithm was arrived at after many hours of experimentation and analysis of its behavior with a wide variety of program material. Many of its final operating constants were arrived at 'empirically' and at this time it represents our best efforts at:

- addressing the tradeoffs associated with processing non-stereo program material, and;
- without the side effects heard in competing products, and;
- and in a manner that creates a subjectively consistent spectral and distortion signature on the air and with real-world program material.

Vorsis® is a registered trademark of Wheatstone Corporation
600 Industrial Drive, New Bern, North Carolina 28562
252-638-7000
www.vorsis.com