True-to-Life Sound Reproduction using

Recursive **A**mbiophonic **C**rosstalk **E**limination (RACE)

by Ralph Glasgal and Robin Miller

November 2006 rev.

Over the last century, the world of stereo matured – in recordings, movies, radio & TV, and personal players. If in the new century you want lifelike realism from your collection of music, movies, and games, then consider the advantages of Ambiophonics. What is Ambiophonics and how does it work? This is your introduction and guide to Ambiophonic 2-channel or Ambiophonic surround sound in your home or studio.

Ambiophonics is so named to suggest that it is the logical next step beyond stereophonics in home theaters, recording studios, or cars in the same way that stereo has replaced some but not all monophonic sound venues (not telephones, AM radio, etc.). Ambiophonics improves reproduction of most stereophonic recordings. PanorAmbiophonic ("PanAmbio") surround does the same for 4.1 and 5.1 music and movies. More about Ambiophonic methodology is available at www.ambiophonics.org.

In essence, one crucial difference between Ambiophonics and stereophonics is the elimination of the acoustic crosstalk, comb filtering and pinna direction finding errors introduced when two channel recordings such as CDs, LPs, or 5.1 DVD/SACDs (all essentially free of such errors as pressed) are reproduced using speakers that form an equilateral triangle with the listener. So doing away with such stereo crosstalk is essential to achieving high fidelity sound reproduction.

The problem with stereo is this: Where in normal binaural hearing just one ray reaches each ear from a sound source, for most sources on a stereo disc, two sound rays, one from each speaker, reach each ear. This is how conventional stereo works, and the deleterious effects are many: These extra rays, or crosstalk, cause high frequency peaks and dips (comb filtering) that confuse the pinna (outer ear), distorting timbre as sound images move from side to center. The crosstalk at lower frequencies also alters the correctly recorded time and level differences between the left and right ears, causing localization to be fuzzy or inaccurate.

Again, when a sound source such as a soloist is centered, speakers, positioned at ±30° toward the sides instead of the front, generate false side pinna localization cues and comb-filtered timbre, altering important solo voices (the reason 5.1 uses a C channel). This 60° front-speaker separation also restricts stage width to that angle. These artifacts (along with rear hall ambience coming wrongly from the front) prevent stereo reproduction from being as compelling as live hearing, e.g. the true-to-life sound of a live concert.

Contrast this with Ambiophonics where the front stage is rendered without pinna confusion, without altering timbre, and with a natural stage up to 180° wide if that has been captured in the recording. Ambiophonics does this by eliminating crosstalk – and by turning stereo inside out – moving the speakers together to the front where they recreate sounds not in between but outside of themselves.

Eliminating crosstalk to improve sound reproduction is a long established principle. However, more recently it has been shown that the only way to have effective crosstalk cancellation is to use a pair of speakers, called an Ambiodipole, that are relatively close together (24° apart or less as viewed from the listening area). Having both speakers directly in front of the listeners means that the crosstalk correction parameters can be assumed to be approximately independent of the shape and size of the listener's head, and this has proved to be true in practice for adults. Of course, personalizing all the cancellers described below is always possible.

There are three different approaches to cancellers used in Ambiophonics. The first is the simple physical barrier – an absorbent vertical wall extending from the listener toward two speakers placed just slightly further apart than the width of the barrier (see Figure 1). Such an arrangement is loudspeaker binaural. Think of Ambiophonics as "virtual headphones" but with the greater comfort and quality of speakers. Compared to earphone listening to stereo or special binaural recordings, the sound is never inside your head, and you can rotate your head without the stage rotating with you. And, since the pinnae are not impacted by the earphones, they can function normally, guaranteeing externalization. (Note: Before investing in more

technological approaches, you can try Ambiophonics simply by erecting a make-shift barrier with speakers on either side.) Barriers normally do not cancel much at frequencies below 400 Hz but this is fine, since little crosstalk is there to cancel. Also, with the speakers so close together, the bass mode response of the room should be easier to control and slap echo is reduced.

Obviously, sitting with your nose at the end of a mattress-like barrier is not for anyone except a determined perfectionist. So over the years there have been attempts to eliminate the barrier, along with the crosstalk, electronically, first using analog and later digital signal processors (DSP). The Ambiophonic team has been working on perfecting crosstalk cancellers for more than a dozen years. With advances in computer power, DSP/PC software, and new inspiration, we now can offer two families of crosstalk cancellers that perform to audiophile standards, are variously affordable, and are user friendly to three major groups of users: home high fidelity enthusiasts, recording/mastering engineers seated at a digital audio workstation (DAW), and computer gurus who prefer to process audio using impulse responses and convolution.

The basic crosstalk canceling technique we have developed and are making available free to the internet community is the Recursive Ambiophonic Crosstalk Eliminator or RACE. Recursive is the operating word. When a signal from the left speaker undesirably reaches the right ear, it must be cancelled at that ear by an inverted, perfectly delayed, slightly lower level replica from the right speaker. But this cancellation signal will also reach the left ear and so it must also be cancelled (2nd order cancellation) by a properly conditioned signal from the left speaker, which signal then also reaches the right ear requiring another round (3rd order) of cancellation, and so on. For a greater tolerance for non-ideal speakers, to avoid frequency response errors, and to enlarge the listening area, this recursive "ping-pong" correction needs to be carried out to inaudibility. We have demonstrated that up to five people can hear the same wide stage even from two small speakers using this method.

Figure 2 is the logical representation of the RACE process showing the feedback loop that keeps the see-saw process going until the digital samples become zero. The process results in correct cancellation for both one-sided and center signals if the assumptions about 1) incremental delay and 2) incremental attenuation are realistic. These are the two adjustable parameters. First is the attenuation at each step of the offside speaker signal as it goes to the wrong ear. For most heads and speaker angles this is –2 to –4dB. Adjusting this parameter is like tweaking a stereo system. This value optimizes the maximum stage width and the size of the listening area. Too little attenuation tends to over-cancel and will exaggerate the stage width or sound a bit like an out of phase stereo system – and may also depress the higher mid-bass frequencies (if bass is unbypassed) where there is no real crosstalk to cancel. Too much attenuation will not completely cancel the crosstalk, so the system will narrow the stage width, but broaden the listening area..

The delay parameter ranges from 60 to 120μs (microseconds) and this value is significant at frequencies below 8kHz, where it is not comparable to the period of sound e.g. 100μs at 10kHz. This delay is determined by the path length difference from a speaker to both ears, and so is a function of the separation of the speakers and their distance to the listener. You want this programmed delay and the actual delay to the other side of your head to match (within reason) and not introduce dips and peaks above 10kHz. These, if not bypassed and thus present, are masked by similar dips and peaks created naturally by the pinna, concert hall, and objects in the room, and are negligible compared to the 1kHz and up comb filtering generated by the 60° stereo triangle. This delay parameter should be as short as possible commensurate with a wide stage so as to move the frequency at which non-minimum phase behaviour and thus erroneous crosstalk cancellation begins at as high a frequency as possible, typically 10kHz or higher.

The assumption made in figure 2 that the attenuation, due to the longer trip past the nose, is constant begins to fall apart below about 500 Hz. The attenuation due to the passage of sound through the air is less at lower frequencies and any part in the process due to the face becomes insignificant. Also at about 400Hz the crosstalk from an Ambiodipole (two speakers typically 10~24° apart) or even from stereo speakers becomes less able to influence localization or stage width because the human head is no longer an effective barrier to sounds with long wavelengths. The difference in level between the two ears at say 250Hz is psycho-acoustically negligible. The time delay is still the same but a delay of say 80μs equals a phase difference of only 7° between the ears at 250Hz and this tiny phase difference aids the brain little in localization. To avoid loss of bass level due to unnecessary processing as well as unnecessary high frequency processing, such

frequencies are bypassed around the RACE process by a digital filter,.and then recombined for reproduction by the Ambiodipole, by subwoofers, and by optional side speakers discussed below.

Setting up a RACE Ambiodipole is easy using the test track downloadable at www.ambiophonics.org. A track with a voice alternating left, right, center allows determining the optimum speaker placement for a given listening position. (Or on a line from between speakers of a given separation (typically 10-24°), listener positions can be found where the stage is as wide as desired without making a noticeable change in the level of a center voice.) When the delay and attenuation parameters are properly adjusted, any timbral or level anomalies should be less than when running the same test in stereo (speakers typically separated 60°). You can make an instant comparison of Ambiophonics to stereo simply by moving closer to the speakers until you are positioned at the apex of an equilateral triangle. With Ambiophonics, four people should be able to sit along the median line and hear a binaurally correct wide front stage.

The RACE algorithm is available in three different forms. It can be ordered as an already installed option in the TACT Audio RCS 2.2-XP – see www.tactaudio.com. For audiophiles who can afford the purchase price, this is an easy way to go. The crossover for subwoofers comes before the RACE but the bass is also always bypassed around the RACE when there is no subwoofer. The delay and the attenuation parameters are adjustable from the remote control and this makes optimizing an Ambiodipole quite easy. Since RACE is in the public domain one can expect that other manufacturers will offer RACE products at various price points. While most speaker pairs with consistent crossover network performance work well as Ambiodipoles, a full range, curved electrostatic product, the Ambiopole, has been especially designed for this application.– see www.soundlab-speakers.com.

For those who are skillful with PCs, inexpensive programs can host the RACE algorithm (in several evolving forms at www.ambiophonics.org and downloadable free). It is best if the soundcard has a SPDIF digital input and output, although analog I/O is also feasible. Figure 3 shows the high-order RACE hosted within AudioMulch. (Additional solutions using hosts including a Nord G2 are at www.electro-music.com.)

Another approach to implementing RACE is to convert the diagram of Figure 2 to an impulse response. That is you feed a needle like pulse into the processor and measure what comes out. You then have what the computer needs to do to any sample presented to it (called convolution) to cancel crosstalk. Such Impulse Responses (IRs), downloadable free at www.ambiophonics.org, are then used by programs called convolvers to perform the RACE process. Convolvers such as Voxengo, X-Volver, and Waves IR-L are VST plug-ins and are readily inserted into host programs like Audio Mulch, Plogue Biddule, Cubase SX, etc.

The Choueiri IR (download at www.ambiophonics.org) is a crosstalk canceling IR not based on RACE. It avoids any bass or any very high frequency response errors and simplifies setup and operation (no adjustments) by those with PC/Mac experience. Still evolving, it is the forerunner of many high-performance IR plug-ins to come that may advance the art of audio convolution considerably.

The above discussion has been limited to the application of RACE to most stereo recordings using a single frontal Ambiodipole during studio monitoring or playback. However, we have discovered that even where only two channel CDs or LPs are involved, one can get impressive improvements in binaural realism by using additional speakers. (A different use of Waves Audio or other concert hall IRs plus a second convolver to derive purely ambience signals for surround speakers is discussed below and elsewhere.) If one puts a second Ambiodipole behind the listening position, and feeds it the same signals, at a slightly reduced level, as the front Ambiodipole, the stage widens toward 180° and there is a greater sense of being there. How does this work? With one Ambiodipole, the attainable stage width is limited to 150° because the sound above 1kHz is from frontal Ambiopole speakers. But for images increasingly away from the middle,, the pinna direction finding mechanism contradicts other interaural level and time cues, and the pinna begin to win at larger angles. However, if a completely contradictory set of high frequency cues comes from behind, the original frontal peaks and valleys are largely leveled mimicking somewhat a side sound source so that the brain now accepts the lower frequency cues at their face value and can localize images at the full ±90° points, particularly for virtual reality panned recordings and movies.

As an alternative to the rear pair or in addition to it, one can place two speakers at the ±90° points (directly left and right) and feed them the unprocessed stereo signals at slightly lower level than the fronts. Again, the stage widens to 180° if fully left or fully right sounds are on the disc. Similar to the old Beveridge

approach, this is another pinna correction device. That is when the side speakers are loudest for full left or right sounds they have a direct line to the ear canal, and thus this pinna pattern helps the brain localize to the far side. The side speakers have little effect on the center stage because the infamous stereo hole-in-the-middle now works to our advantage. As happens in 5.1 speaker layouts, it is possible with some recordings to hear a combing effect (a sort of raspiness) when the side speakers are operating. Modifying their level, delay, or position can usually eliminate this effect. (We suggest you to try extra speaker pairs after you have become accustomed to front Ambiodipole RACE.)

In order to restore the ambience that in most 2 channel stereo recordings is weak (because it would sound wrong, coming as it does only from the front), a great Ambiophonic system includes surround speakers whose sole purpose is to restore ambience – mimicking concert hall walls and a psycho-acoustically correct ambient sound field that can enhance the front stage that RACE has created. To have a true-to-life concert hall experience at home, the early reflections and the later ambient tails would have to come from the same directions as they would have if you were in the hall where the recording was made, at least, within reason. Using Audio Mulch, Waves IR-L and the set of hall IRs provided by Waves, you can generate as many diverse uncorrelated ambience signals as you like for almost as many speakers as you have room for. Figure 4 shows ambience convolution for 16 speakers using Waves IR-L. Figure 5 shows high order RACE along with convolved ambience for a total of 18 speakers, depending upon PC capability. See www.ambiophonics.org for the Audio Mulch configurations that make this possible. Of course the PC must have an output for each speaker, and so is limited by how many sound cards you can drive. We have driven 14 surround speakers from one computer using RME cards (ADAT-based), but this approach requires computer savvy. It is anticipated that future home theater processor products will include hall IRs and built-in convolvers.

So much for enhancement of 2-channel stereo recordings – what about multi-channel 5.1 surround music and movies? The same simple tools at www.ambiophonics.org can be used double, feeding a second Ambiodipole in back to reproduce the SL and SR channels (see Figure 6). Set the player for "no center (speaker)" so that the C-channel is mixed into the RACE, where it creates a perfect center image. The front Ambiodipole yields a frontal movie sound stage of 180 degrees and the rear Ambiodipole provides the rear half circle for easy localization of movie sound effects such as gunshots, flyover, growls and other mayhem.
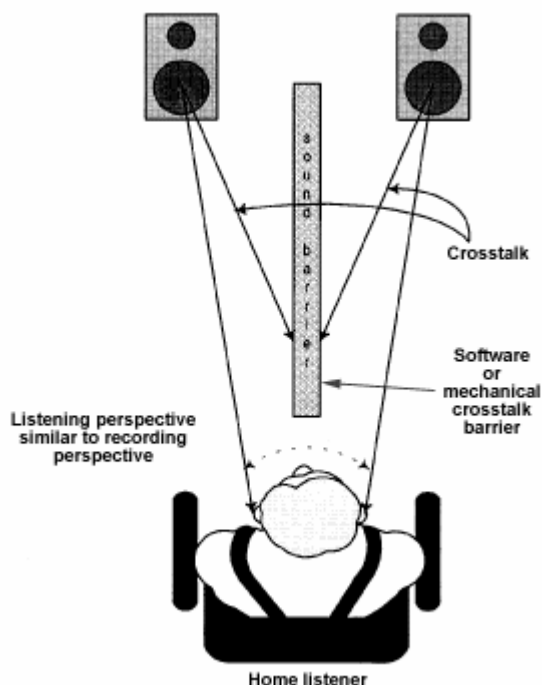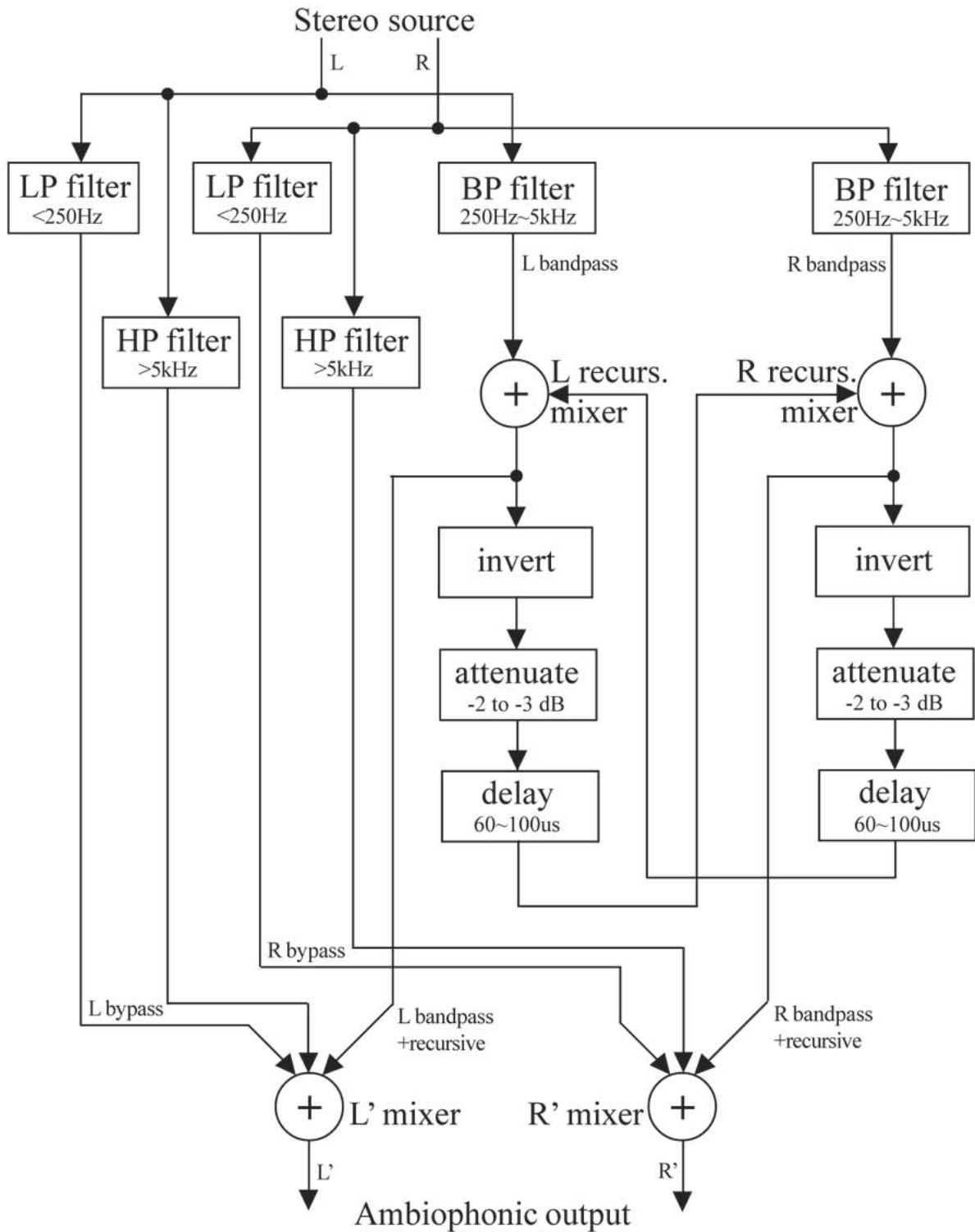
Figures

Fig.1 – illustrates the simplest form of crosstalk cancellation using a physical barrier, now replaced by Digital Signal Processing (DSP).

Recursive Ambiophonic Crosstalk Eliminator (RACE)

Fig.2 – R.A.C.E. (recursive Ambiophonic crosstalk eliminator) DSP block diagram shows side-chain that creates an infinite series of inverted, attenuated, and delayed cancellation signals for a 2-channel stereo input. RACE equations are in Appendix A below.
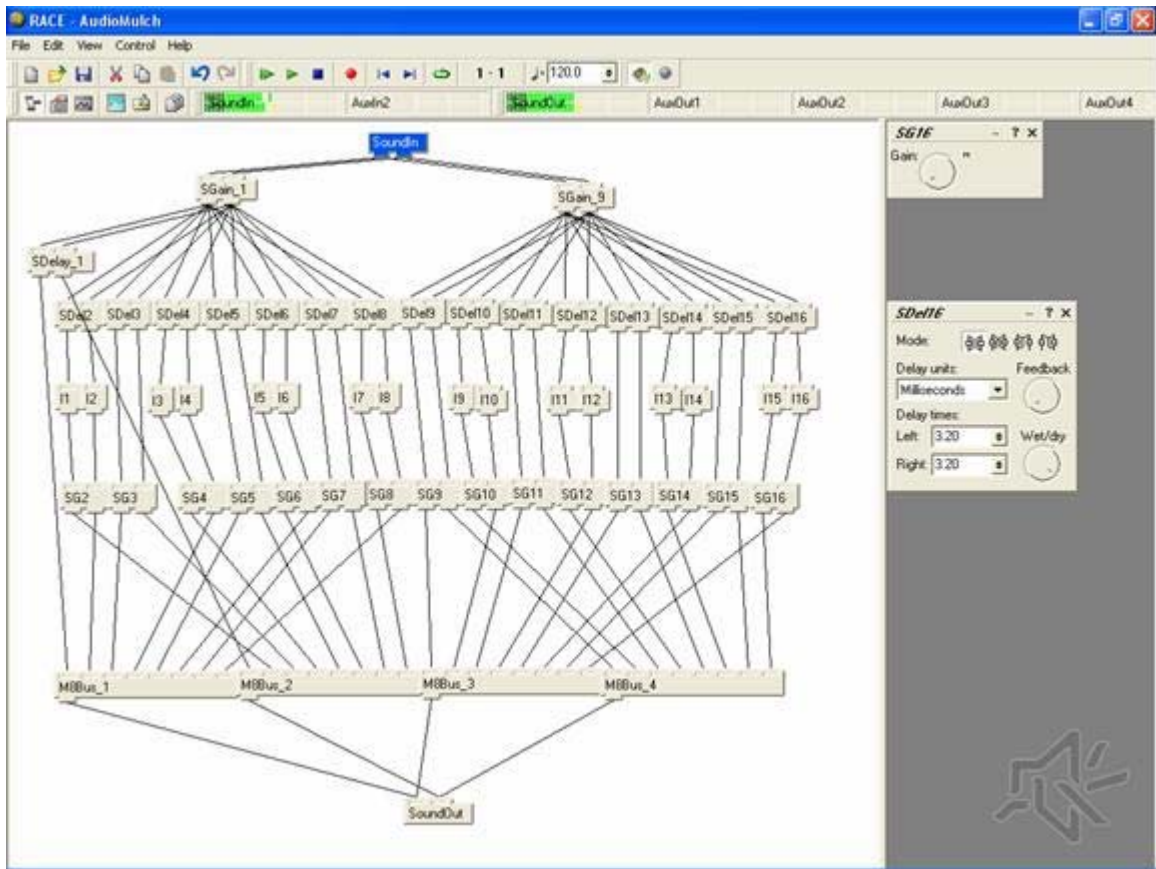
Fig.3 – high-order R.A.C.E. hosted in AudioMulch using downloadable software available at www.ambiophonics.org.
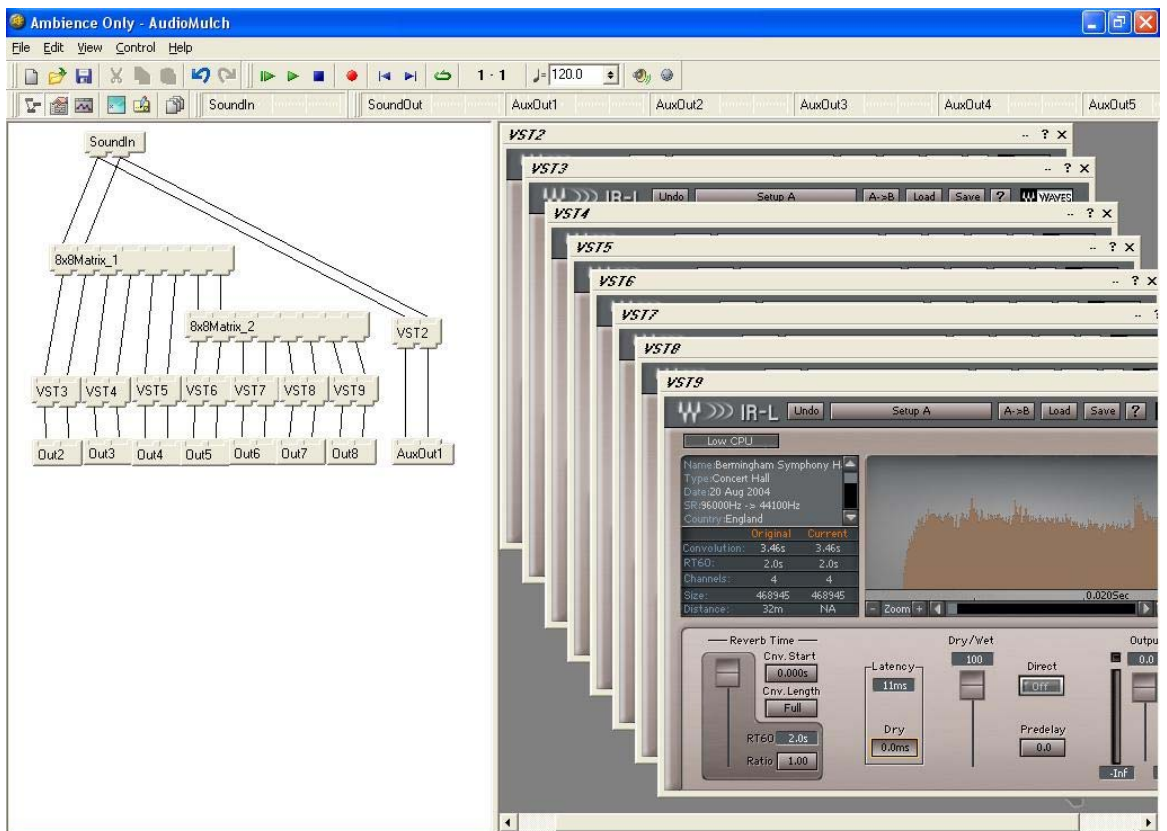

Fig.4 – Ambience by Impulse Response convolution shows 9 incidences of Waves IR-L within AudioMulch feeding 16 speakers.
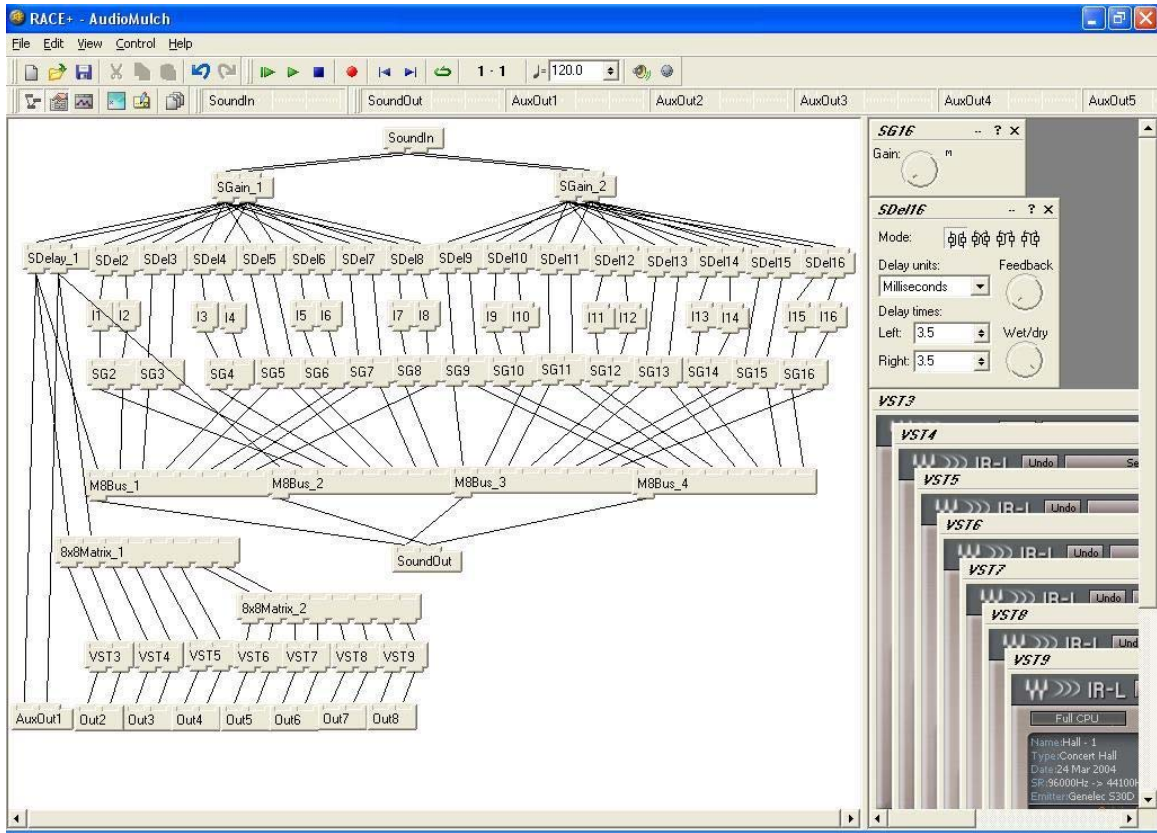
Fig.5 – Combined high order RACE (Fig.3) and AMBIENCE (Fig.4) feeding 18 speakers depending upon PC capability.
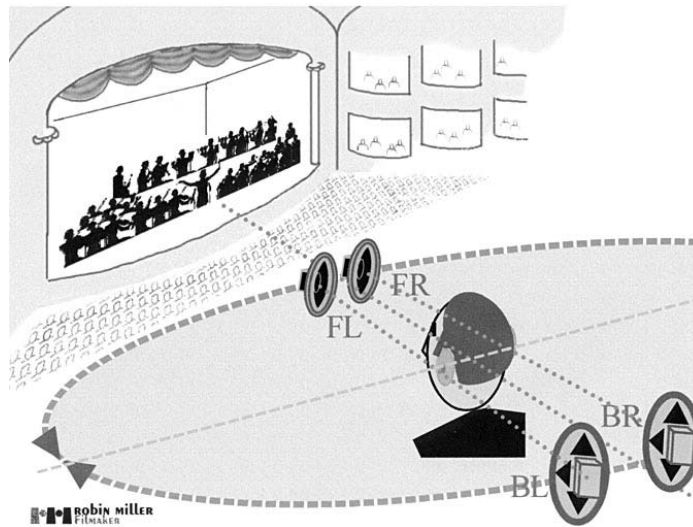

Fig.6 – PanAmbio surround reproduction of multi-channel 5.1 movies and music provides a second rear stage for surround signals, whether antiphonal voices or 2D ambience reflections. The C-channel is mixed into the front speakers to produce a center image.

———

APPENDIX A – R.A.C.E. EQUATIONS

A = unit of attenuation, 2A = 2 increments of attenuation, etc.

D = unit of delay, 2D = 2 units of delay, etc.

RACE output signals (for a right-only input, one PCM sample) $L_{in} = 0$, $R_{in} = R_{out} = R_{0D0A}$

$R_{out} = R_{0D0A} + R_{2D2A} + R_{4D4A} + R_{6D6A} + R_{8D8A} + ...\ =\ {}_{n=0}\Sigma^{\infty}R_{2nD2nA}$

$L_{out} = - R_{1D1A} - R_{3D3A} - R_{5D5A} - R_{7D7A} - R_{9D9A} - ... = -\ {}_{n=0}\Sigma^{\infty}R_{(2n+1)D(2n+1)A}$

Similarly for a Left-only input (channels reversed).

At the ears = RACE output signals + acoustic crosstalk for an Ambiodipole:

For a R-only input, sound reaching the Right ear = $R_{0D0A} - R_{2D2A} + R_{2D2A} - R_{4D4A} + R_{4D4A} - R_{6D6A} + R_{6D6A}...$

i.e. Right <u>out</u> + Left <u>out</u> further delayed & attenuated acoustically by 1D1A = R_{0D0A} = R_{in} (the R-only input)

Sound reaching the Left ear

$L_{ear} = L_{out} + R_{out}$ incremented by 1D1A $\ = - R_{1D1A} + R_{1D1A} - R_{3D3A} + R_{3D3A} - R_{5D5A} + R_{5D5A} -...\ = 0$ (silent)

Similarly for a Left-only input, L_{ear} = L-only input, while R_{ear} = 0 (silent), so crosstalk is cancelled at the ears.

For a central image, the common (monaural equivalent) signal $M_{in} = L_{in} = R_{in} = M_{0D0A} = L_{0D0A} = R_{0D0A}$

Sounds reaching the ears

$R_{ear} = R_{0D0A} - R_{1D1A} + R_{2D2A} - R_{3D3A} + R_{4D4A} - R_{5D5A} + R_{1D1A} - R_{2D2A} + R_{3D3A} - R_{4D4A} + R_{5D5A} -...$

$= R_{0D0A}$ (i.e. the common M_{in} signal)

$L_{ear} = R_{0D0A} - R_{1D1A} + R_{2D2A} - R_{3D3A} + R_{4D4A} - R_{5D5A} + R_{1D1A} - R_{2D2A} + R_{3D3A} - R_{4D4A} + R_{5D5A} -...$

$= R_{0D0A} = L_{0D0A} = R_{in} = L_{in} = M_{0D0A}$ (i.e. again the central sound image $M_{in,}$ thus producing a central image)

Authors

Ralph Glasgal, B.E.P., M.S.E.E. (IEEE, AES) a Cornell University Engineering Physicist and Electronics Engineer, was awarded a patent for a stereo dimension control and designed recording equipment for RCA and high fidelity components for Fisher Radio. He has authored many articles for magazines including Stereophile and The Audiophile Voice. He is the founder of the Ambiophonics Institute and the co-author of Ambiophonics-Beyond Surround Sound to Virtual Sonic Reality. Glasgal Island in the Antarctic was named for him in recognition of his research in the south polar region and he is a noted authority on wide-area networking and data communications.

Robert E. (Robin) Miller III, BSEE, AES, SMPTE is a musician, an orchestrator, and a filmmaker recognized by 52 awards including The Peabody. He has more than 40 years experience in music recording and mixing films and television specials. As president and CTO of FilmakerTechnology, he develops advanced entertainment technologies and has a Patent Pending on an advanced stereo and 5.1-compatible 3D (full sphere) reproduction system.