AMBI@Home - The search for extra-frontal intelligence

Linux Audio Conference 2008, Cologne

Jörn NETTINGSMEIER
Sound and event engineering
Lortzingstr. 11
D-45128 Essen
nettings@apache.org

Abstract

This paper describes the setup and calibration of a Do-It-Yourself Ambisonic surround listening rig made from comparatively cheap components. The rig comprises a Linux audio workstation with JACK, Ardour2, AmbDec, and the JACE convolution engine. The DRC room correction software is used to compensate for room and speaker deficiencies. It is shown that a determined amateur can obtain very good results using only free software and commodity hardware in a standard domestic environment, and that dedicated Ambisonic playback systems based on such components might eventually be commercially feasible for a limited market of enthusiasts.

Keywords

Ambisonics, surround sound, room correction, home theatre

1 Introduction

Ambisonics is a surround sound methodology developed in the 1970s by Michael A. Gerzon et al. [Ger1974]. It is based on the three-dimensional deconstruction of a given sound field using spherical harmonics, which can then be reconstructed on different speaker layouts with an appropriate matrix of decoding coefficients.

The most compelling features of Ambisonics are the decoupling of transmission and rendering formats (which means that the producer of surround content need not take into account the consumer's setup), and its scalability from mono to full 3D (or *periphony)* in various *orders* of directional precision. Even more important for home use, the sonic result degrades gracefully when the number of available speakers or transmission channels is limited.[1]

The transmission representation of Ambisonics is called *B-format*. In its most widely used *first-order* form, it contains an omnidirectional pressure signal W and three figure-of-eight side difference or velocity signals X, Y and Z (left - right, front - back and up - down).[2]

Together these four channels comprise a complete three-dimensional representation of a sound field in one point, which compares favourably to the six transmission channels used by Dolby 5.1 for planar-only surround.

In addition to its technical merits, Ambisonics fits well into the free software ecosystem: based on a rigourous mathematical foundation, it offers excellent introductory documentation, a wealth of advanced scientific papers and a number of high-quality open-source implementations. But most importantly, the technology itself is freely available, since all relevant patents have now expired.

1.1 Localisaton cues and shelf filters

The Ambisonic enthusiast should have a basic understanding of the different mechanisms of

[1] If necessary, an Ambisonic signal can be replayed in mono without phase cancellation effects.

[2] Readers familiar with stereophonic miking techniques will be reminded of a Blumlein pair or an MS setup, extended to three dimensions.

sound localisation. Below about 700 Hz, humans can only rely on phase difference at the ears - amplitude differences will be negligible at those wavelengths, because the sound is diffracted around the head without significant shading effects.

As the wavelength approaches twice the distance between the ears, phase cues become ambiguous. Consequentially, in the 700 Hz to 5 kHz range, the primary directional cue shifts to loudness difference [Ger1974].

So-called *classic* ambisonic decoders make use of this fact by using two sets of decoding coefficients, to maximise the *velocity vector* $\mathbf{r}_V$ (for phase information) below 700 Hz and the *energy vector*[3] $\mathbf{r}_E$ (for loudness information) at higher frequencies respectively.

The shelf filters used to separate those bands must be phase-matched [Lee2007]. Since this is non-trivial, some software decoders do not implement them. Two-band shelf-filtered operation will yield superiour quality in small-scale setups and should be preferred over simpler designs.

1.2 Virtual sound sources

When tweaking and testing an Ambisonic setup, it is important to know that virtual sources are never rendered by only one speaker. In fact, in most home setups, all speakers will contribute to any one virtual point source - those close to the source with in-phase signals of varying intensity, those opposite with out-of-phase signals.

The Ambisonic novice might be confused by the fact that meter readings of the speaker feeds will provide no useful clues as to the location of a sound, and even the B-format readings take some experience to become useful.

In a properly tuned Ambisonic playback system, the speaker positions are inaudible and virtual sources will sound the same whether they are on or between speakers. This allows for perfectly smooth panning without sounds being pinned to speaker locations. The tradeoff is a slightly less focused and sometimes spectrum-dependent image.

1.3 Near-field effect

Basic Ambisonic recording systems assume that all incoming sounds are plane waves, which is equivalent to their distance being infinite.

Naive decoders make the same assumption for the speakers of the reproduction rig.
For real-life sources and speakers, the distances will be finite (often very small) and the actually recorded and reproduced wave fronts will be curved. Since the X,Y and Z signals and all speakers are directional, this results in a boost of low frequencies. Users of directional microphones will recognize this as the *proximity* or *near-field* effect.[4]

If this effect is not compensated for during reconstruction, the resulting sound will contain way too much bass. Since the amount and threshold frequency of the boost are a function of the order of the system [Dan2006],[5] first-order rigs are somewhat forgiving for most material (the effective gain will be 6 dB/oct with +3 dB at about 50 Hz). However, organ music with 32 ft stops and similar material will suffer.

At second or higher order, decodes without near-field compensation will quickly become intolerable.

1.4 Speakers

Successful Ambisonic reconstruction requires precisely matched levels and a uniform phase response of all speakers. Therefore, mixing speaker brands or worse yet, entirely different construction principles, will almost certainly produce poor results.[6]

Depending on the budget, a number of speaker layouts are possible. Generally, image stability and

[3] [LaH2007] note that the term *energy vector* is misleading, as energy is a scalar quantity. It might be clearer to think of the *energy gradient* instead.

[4] A good explanation is in [WP2007]. Note that the dominance of the inverse-square law of the amplitude over phase differences in that article is equivalent to the waveform curvature mentioned here.

[5] The boost has a slope of $m \cdot 6$ *dB/oct* with infinite(!) gain at 0 Hz, where m is the order of the Ambisonic system.

[6] Although it might be interesting to try whether aggressive phase response correction can reconcile different speaker models and provide at least satisfactory results, to allow consumers to re-use and extend existing speaker sets.

size of sweet spot increase with the number of available speakers.

The minimum number required for stable planar surround is four, aligned in a square or rectangle. Benjamin, Lee and Heller have studied rectangular rigs in detail and offer setup recommendations based on thorough listening tests [BLaH2006].

If 3D reproduction is desired, the minimum stable setup consists of eight speakers arranged in a cube[7].

For optimal results, precise placement of speakers is important. When planning an Ambisonic rig, care should be taken to position the speakers on the nodes of a regular polygon or (in the 3D case) polyhedron as much as possible.

Oblong shapes (such as the rectangle examined by [BLaH2006]) have proven practical if the program material is mostly concentrated on a frontal sound stage, but they do have worse localisation to the sides.

If an ideal regular layout is not possible, slight variations in distance are to be preferred over incorrect azimuth angles, since they can be trivially corrected with delay without affecting the decoding matrix coefficients. In any case, each speaker should have a precisely diametrical opposite.

For irregular configurations, there exists no straighforward algorithm to determine the decoding matrix coefficients for $\mathbf{r}_V$ and $\mathbf{r}_E$ optimisation.[8]

It is common wisdom that, room size permitting, speakers should be placed well away from walls[9], so that the first reflections are sufficiently late to be perceived distinctly and not as coloration interference. However, if digital room correction is to be used, it might be advisable to locate the speakers close to walls or even corners, because wall reflections can be corrected the more easily the tighter they are coupled with the speaker [Sbr 2005-2]. Earlier reflections can be compensated with shorter filters, which demand less computing power.

2 Phase 0: The gear

The central component of the prospective Ambisonic playback system is the author's audio workstation, an Athlon64 4000+ with 2 GB of RAM. Even with instrumentation scopes and other CPU-hungry helpers, the described setup of six real-time convolutions and a running Ardour instance rarely exceeds 25% CPU usage, so it should be perfectly feasible to implement a similar system on a shoebox PC with passive or at least very quiet cooling.

The audio interface is an RME Digi 9652 connected to an external Behringer ADA 8000 eight channel AD/DA converter and mic preamp. The converter drives six active Tannoy 5a monitors. A Behringer ECM 8000 omni-directional instrumentation microphone is used for measurements.

The hub of the software stack is a JACK daemon [Dav2006], which enables real-time audio data exchange between the numerous other components.

Ardour2 [Dav2007], a comprehensive digital audio workstation, is used to play back the B-format recordings. It is configured to use a 4-channel master bus with all panning plugins disabled, to make sure the B-format is passed through without errors. Ardour is jacked into AmbDec [Adr2007], an Ambisonics decoder which converts the B-format input to suitable speaker feeds. Each of the six outputs of AmbDec is then patched into a real-time convolution engine (JACE, [Adr2007-2]) that applies pre-computed correction filters before the signal is fed to the speakers.

During the calibration phase, Aliki [Adr2006] was used to record and compute the impulse responses. The filter kernels were computed by DRC [Sbr2005], and the JAPA analyser [Adr2007-3] provided quick frequency response checks in real time.

[7] In theory, the minimum number of speakers is equal to the number of B-Format signals. Gerzon has described triangular (2D) and tetrahedrical (3D) setups, but these are mathematical constructs rather than actually usable configurations, as they provide poor image stability. However, these configurations do have their use in binaural rendering.

[8] Adriaensen provides a matrix for the highly irregular ITU 5.1 setup that was obtained by genetic search [Adr2007] (compare Lee and Heller [LaH2007] for a detailed examination); however, applying such methods to arbitrary layouts might be beyond the average Ambisonics amateur.

[9] unless, of course, one uses small speakers that would produce insufficient bass without the help of a rear wall or corner

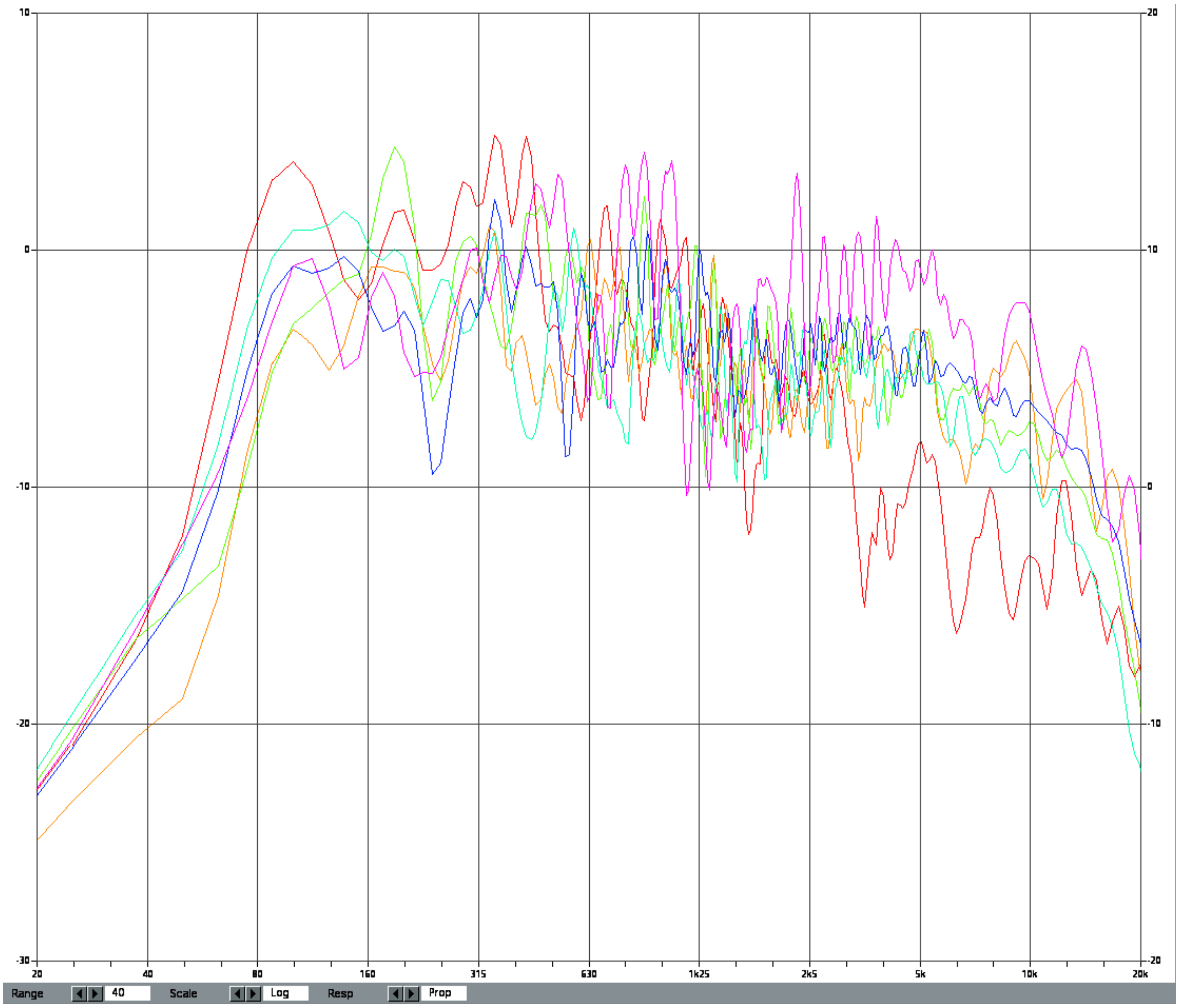

Illustration 1: Speaker frequency responses without correction. Note the obvious misbehaviour of the front speaker (red graph). Measured with pink noise, obtained using Japa [Adr2007-3] set to Bark scale, medium resolution, "noise" setting.

3 Phase 1: Speaker setup

As the budget was limited to six speakers at reasonable quality, it was decided to forego full periphony and install a planar (2D-only) hexagon instead[10]. However, the methods described below will also apply to 3D rigs or other planar speaker layouts[11].

The listening room was quite small, its dimensions being 3.75 x 3.10 x 2.50 m, with a window in the back, a large bookshelf covering the entire left wall, and a waist-high bookshelf on the right wall. For practical reasons, the hexagon was initially orientated with one speaker to the front.

After the speakers had been placed at preliminary positions and their locations roughly measured, the central sweet spot was defined by a marker on the floor and the instrumentation microphone was placed above it at ear level[12]. A laser range finder and a cheap laser angle gauge mounted on a camera tripod were used to fine-tune the speaker positions.

[10]influenced by [BlaH 2006], who found the hexagon to be superiour to square and rectangular layouts in the general case (i.e. not exclusively frontal sound stages)

[11]It should be noted that slightly different shelf filters must be used for periphonic setups [Lee2007].

[12]To avoid directional errors due to diffraction and shading effects of the capsule housing, the microphone was pointed straight downwards, which should give identical response for all angles on the horizontal plane.

The azimuths were almost perfect, with the exception of the left-front one, which had to be placed at 55° instead of 60°. The distances (measured from the cone of the tweeter) varied from 1.56m for the front and back speakers to around 1.35m for the side speakers, again with the exception of the left-front one, which had to be placed at 1.62m. Speaker heights varied from -5cm on the left to +5cm on the right.

The speaker distances were then entered into AmbDec, using the unmodified matrix coefficients of the example hexagon template. As of version 0.1.0, AmbDec does not yet take azimuth and elevation into account, but it does provide optional distance compensation of both gain and delay based on the values provided. Delay compensation was enabled, gain matching was done manually (see below). Near-field compensation was activated on the outputs.

With the basic setup completed, a test signal [Hel2007] was played to check for obvious wiring mistakes or phase inversions.

A first listening test with a Soundfield recording of Benjamin Britten's Simple Symphony op.4 proved quite enjoyable but showed significant imaging distortion due to level mismatches. It was also found that the front speaker (which had undergone factory repair earlier) had a severely degraded frequency response (see illustration 1).

4 Phase 2: Speaker and room correction

4.1 IR measurements

In order to correct the problems encountered in the first listening session, two approaches were considered: to correct only the speakers, or (to a certain extent) the entire system of speakers and room. The first approach would have required free-field measurements out in the open, as no sufficiently large room (where the early reflections hit well after the direct sound) was available.

Thus, for practical reasons and due to time restraints, it was decided to measure the speakers at their final locations, combining the effects of speaker and room acoustics.

> *For best results, speakers should be measured and corrected independently of room acoustics, so that the problems of each realm can be identified and treated separately. The speaker phase response can then be tackled with aggressive non-minimum phase filters that would produce unacceptable pre-ringing artefacts outside the sweet spot when used for room correction. Room deficiencies can be cured by an effective combination of acoustic means and DRC.*
> *This implies two separate convolution stages during the calibration work, which is easily handled by today's middle-of-the-road CPUs. Later, the two stages can be folded into one to reduce CPU load.*

The frequency and phase responses of the speakers were taken with the sine-sweep technique introduced by Angelo Farina [Far2000], using Fons Adriaensen's Aliki tool. It is based on the idea that a long sine sweep can be deconvolved into (and used instead of) an impulse for IR measurements. A sweep can contain much more energy than a pulse, which greatly improves the signal-to-noise ratio, and it identifies and optionally ignores non-linear distortion, since the deconvolution step projects distortion artefacts of the resulting impulse into negative time, where it can easily be removed [Adr2006] prior to further processing.

It was found that initial measurements were conducted at too high a level. While speaker distortion and rattling noises from furniture items should be mostly irrelevant due to the windowing nature of the sine sweep method mentioned above, it seems that some room modes were actuated in a rather extreme way. This led to overcompensation in the later room correction stages and subsequent overloading of the woofers, so the first set of IRs had to be discarded.

> *Generally, it seems advisable to measure at or slightly below the desired listening level - the signal-to-noise ratio and disturbance suppression obtained by Farina's method is exceptional, and there is no need to go to the limits of the equipment, as one would do when using traditional pistol shot or ballon pop impulses.*

4.2 Digital room correction

After deconvolution, the impulse responses of the six speakers were fed to Dennis Sbragion's DRC program [Sbr2005]. It performs a number of calculations on the IRs in order to generate an inverse filter that can then be convolved onto the speaker feeds in real-time to counteract room and speaker deficiencies.

The first processing stage of DRC normalises and trims the IRs, which is rather unfortunate for the purpose at hand, since information about speaker distance and level mismatch is lost. It would be a welcome addition if the preprocessing were to leave the IRs alone, so that necessary delays and make-up gains were included in the final filters.

Next, DRC searches for deep, narrow troughs in the frequency response and eliminates them. This avoids excessive boosts in the final correction filter, which might damage the equipment. As troughs are usually caused by destructive interference that will cancel the frequency in question regardless of its level, boosts will be ineffective and put needless strain on the equipment. Moreover, the ear is rather more tolerant regarding dips as it is towards peaks [Ger1991].

In addition to frequency response correction, digital filtering can also improve the phase response of a speaker. Since the audibility of relative phase (especially in reverberant recordings) is doubtful except in the case of transients [Esp2002], phase issues are often considered to be of secondary importance. However, a matching phase response across all speakers is paramount for good Ambisonic imaging. A detailed look at phase correction will be the subject of an extended version of this paper.

Finally, all room effects that cannot be properly inverted must be eliminated from the IR, concentrating on the so-called minimum phase components.

Usually, these will be early reflections that retain a fixed phase relationship to the direct signal, particularly low-frequency ones. High frequency problems could be treated as well, but only at the cost of a dramatically shrunken sweet spot and extreme sound deterioration elsewhere.[13] Diffuse reverb and echoes cannot be undone by digital room correction.

It should be emphasized that digital room correction is no replacement for proper acoustic room treatment, but a complement: it works best in the low end, where acoustic countermeasures are costly and invasive, and acoustic tuning is straightforward for the treble range, where digital correction becomes infeasible.

DRC offers a wealth of knobs to fine-tune the results. It was decided to go for gentle correction with a large sweet spot and a psychacoustic weighting of frequency bands, using the ERB preset provided with DRC. It was adapted to a sampling rate of 48k but otherwise used unchanged.

After some computation, DRC yielded six filter kernels which were then loaded into the JACE convolver engine.

5 Phase 3: Level matching

Level checking was performed with pink noise, by patching the instrumentation microphone into a suitably high-res RMS meter, and verified by ear using a monophonic music signal routed to each speaker in turn. Additionally, a monophonic source was moved around 360° with an Ambisonic panner plugin [Adr2006-2] patched into Ardour, and checked for obvious loudness changes.

The matching itself was done in the analog domain, using the speakers' gain controls.

6 Phase 4: Reflection

The initial decision to orientate the hexagon with one speaker to the front was found to be sub-optimal. While it had a strong initial appeal and gave a spectacular sense of immersion, the location of orchestral instruments was weak.

With a rotator plugin set to -90° (i.e. a right-angle turn to the right), a cross-check with a two-

[13] The compensation of non-minimum phase errors requires acausal filters that will lead to very audible and generally unacceptable pre-echoes outside of the sweet spot [Ger1991]. Additionally, acausal filters will introduce significant delay, which poses synchronisation problems in A/V applications.

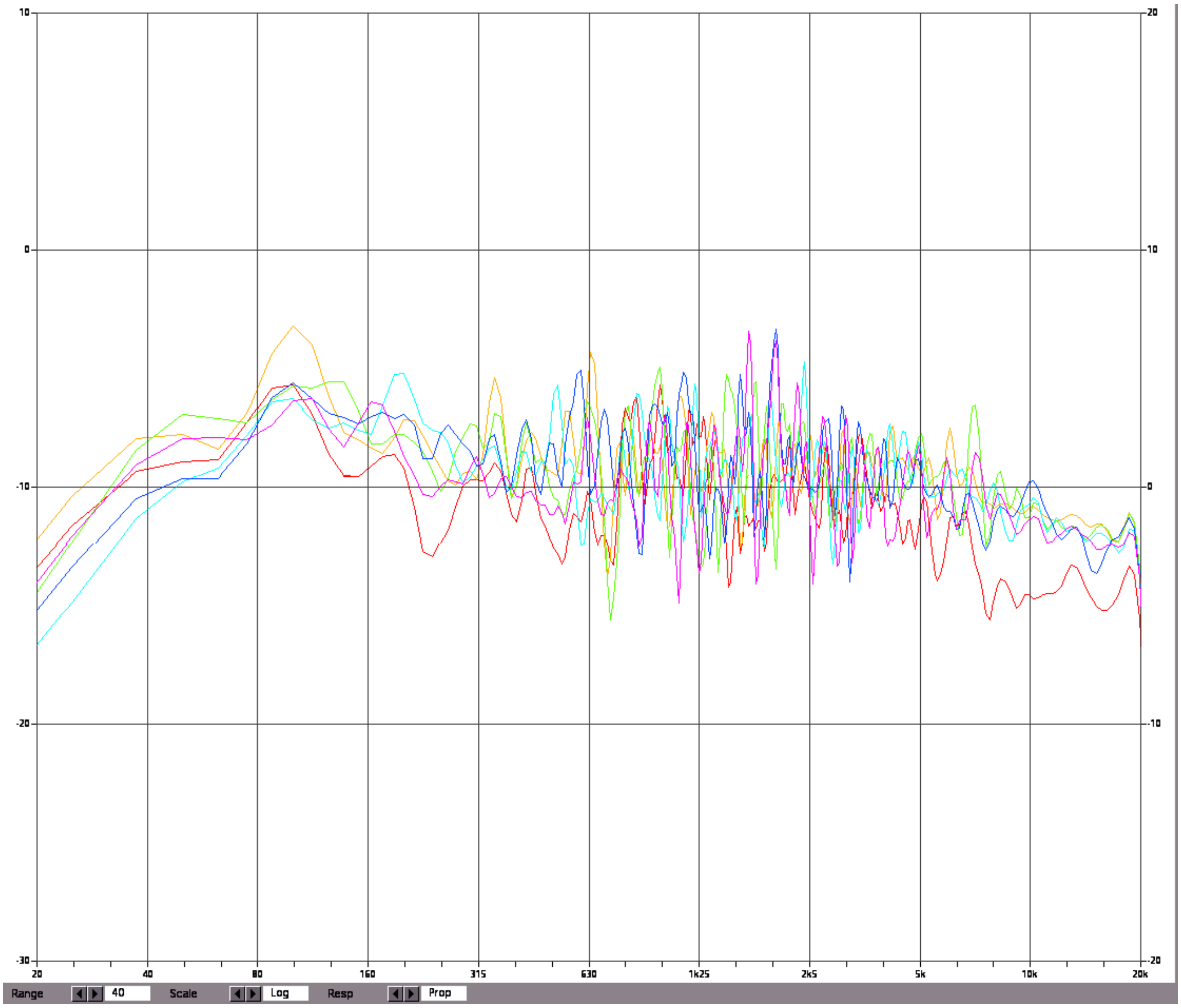

Illustration 2: Frequency responses after correction and level matching. Note the significantly reduced spread and the extended bass range. Front speaker performance is still somewhat unsatisfactory. Same Japa settings as before.

in-front setup could be performed without any shuffling of gear. This setup was better suited to traditional music recordings, as it offered much more stability and precision in the frontal sound stage with only a slight degradation of spaciousness and less coloration.

Hence, the two-in-front orientation was chosen as the final configuration and the listening chairs were turned to the right.

It is quite likely that this outcome was caused by peculiarities in the rig rather than being a general property of the hexagon: the one-in-front layout was slighly elongated, whereas the two-in-front one was squashed, and the (previously repaired) front speaker performed unsatisfactorily even after compensation (see illustration 2).

Moreover, the initial left-to-right acoustic asymmetry caused by the book shelves was traded in for front-to-back asymmetry, to which the human ear is a lot less sensitive, especially when the program material is focused on a frontal sound stage.

Despite the small size of the rig and the listening room, the sweet spot turned out large enough for two people with only slight degradation of imaging. Highly reverberant recordings were found to be most effective, which seems to be due to the dry room acoustics.

The application of digital room correction brought significant improvements even without manual tuning of the DRC algorithm parameters (see illustration 2).

The improvements in the bass range were most evident: almost an entire octave of linear response was gained, and the low end sounded more precise and much tighter. This is consistent with Gerzon's observation that proper equalisation with a filter longer than the room's decay time can significantly reduce room reverberation at low frequencies [Ger1991].

The combined area of six six-inch woofers (roughly equivalent to a 15” speaker) was able to reproduce convincing double basses and timpani without the aid of a dedicated subwoofer.

Moreover, the precision of Ambisonic localisation was clearly improved, images were more focused and phantom sources were (at least in the sweet spot) mostly eliminated.

Some remaining coloration issues will be subject to further experiments. More detailed phase measurements of both room and speakers will be needed in order to further optimise the DRC parameters, and IR-based graphs will have to complement the rather vague noise plots used so far.

It will also be interesting to test whether a more aggressive isolated phase response correction of the speakers with non-minimum-phase filters (which necessitate free-field measurements) can increase localisation and image stability without adverse effects. Finally, the effect of artificial reverb to improve localisation in a dry domestic environment could be examined [Ger1974].

The clear improvements notwithstanding, the listening tests have also indicated that ultimately, the achievable *tonal* quality (at least outside the extended bass range) is still very much bounded by the quality of the speakers and amplifiers. No DSP of this world will ever reduce harmonic distortion, speed up slew rates, double the physical size of a woofer, or tame over-eager or misaligned limiters.

7 Conclusion

It has been shown that even on a limited budget, good results can be achieved with a home Ambisonic setup based on a general-purpose Linux audio PC and affordable speakers. Digital room correction proved useful in matching differing speaker characteristics and in working around the architectural and acoustic constraints typically faced in a domestic environment.

To become commercially feasible for a limited market of audio enthusiasts, the setup procedure would either have to be automated or provided as a service by prospective vendors. Two important missing links are a user-friendly player interface with media library (existing solutions could easily be extended to support the .amb B-Format file type[14]) and ready-to-use renderers that allow existing stereo and 5.1 content to be played back on Ambi rigs.

Further studies might examine setups made from one or more cheap surround speaker sets with satellites and subwoofers. If found usable, such systems could open Ambisonic surround to people on even tighter budgets.

In gaming engines, Ambisonics has recently gained some foothold for internal uses [Del2007], and it is to be hoped that it will eventually make its way into the homes of consumers as a playback format.

[14] Applications using the latest verson of libsndfile [Cas2007] will gain .amb support with little extra effort. The author uses a version of Mplayer that already supports playback of 4-channel .wavs over jack, but since the JACK ports are removed and re-created for every new file loaded (which requires manual re-patching), it is not suitable for end-users.

8 References

8.1 Papers

[Adr2006] Fons Adriaensen, Acoustical Impulse Response Measurement with ALIKI, http://kokkinizita.net/papers/aliki.pdf, 2006

[BLaH2006] Eric M. Benjamin, Richard Lee and Aaron J. Heller, Localization in Horizontal-Only Ambisonic Systems, http://www.ai.sri.com/pubs/files/1374.pdf, 2006

[Dan2006] Jerôme Daniel, Spatial Encoding Including Near-Field Effect, http://gyronymo.free.fr/audio3D/publications/NFC-HOA_AES23.pps, 2007

[Del2007] Etienne Deleflie, Interview with Simon Goodwin, http://blog.ambisonia.com/2007/08/30/interview-with-simon-goodwin-of-codemasters-on-the-ps3-game-dirt-and-ambisonics/, 2007

[Esp2002] Arnold Esper, Hörbarkeit mikrozeitlicher Strukturen im Musiksignal, Verlag Peter Lang, Frankfurt 2002

[Far2000] Angelo Farina, Simultaneous measurement of impulse response and distortion with a swept-sine technique, AES Preprint, http://pcfarina.eng.unipr.it/Public/Papers/134-AES00.PDF, 2000

[Ger1974] Michael A. Gerzon, Surround Sound Psychoacoustics, Wireless World, http://www.audiosignal.co.uk/Resources/Surround_sound_psychoacoustics_A4.pdf, 1974

[Ger1991] Michael A. Gerzon, Digital Room Equalisation, Studio Sound, http://www.audiosignal.co.uk/Resources/Digital_room_equalisation_A4.pdf, 1991

[LaH2007] Richard Lee and Aaron J. Heller, Ambisonic Localisation, Part 2, http://www.ai.sri.com/ajh/ambisonics/AmbiLoc2.pdf, 2007

[Lee2007] Richard Lee, SHELF FILTERS for Ambisonic Decoders, http://www.ambisonia.com/Members/ricardo/shelfs.zip , 2007

[Sbr2005] Dennis Sbragion, DRC documentation, http://drc-fir.sourceforge.net/doc/drc.html, 2005

[WP:2007] Wikipedia, Proximity effect (audio) http://en.wikipedia.org/wiki/Proximity_effect_%28audio%29

8.2 Software

[Adr2006-2] Fons Adriaensen, AMB ambisonic LADSPA plugins, version 0.3.0 http://kokkinizita.net/linuxaudio/downloads/AMB-plugins-0.3.0.tar.bz2, 2006

[Adr2007] Fons Adriaensen, AmbDec, an Ambisonic decoder for first and second order, version 0.1.0, http://kokkinizita.net/linuxaudio/downloads/AmbDec-0.1.0.tar.bz2, 2007

[Adr2007-2] Fons Adriaensen, JACE Convolution Engine, version 0.2.0, http://kokkinizita.net/linuxaudio/downloads/jace-0.2.0.tar.bz2, 2007

[Adr2007-3] Fons Adriaensen, Japa, version 0.2.1, http://kokkinizita.net/linuxaudio/downloads/japa-0.2.1.tar.bz2, 2007

[Cas2007] Eric de Castro Lopo, libsndfile, http://mega-nerd.com/libsndfile/, 2007

[Dav2006] Paul Davis et al., The JACK Audio Connection Kit, http://jackaudio.org, 2006

[Dav2007] Paul Davis et al., Ardour Digital Audio Workstation, version 2.1, http://ardour.org/, 2007

[Hel2007] Aaron J. Heller, Eight Directions (ambisonic test signal), http://www.ambisonia.com/Members/ajh/ambisonicfile.2007-05-29.7251031563/, 2007

[Sbr2005] Dennis Sbragion, DRC: Digital Room Correction, version 2.6.2, http://drc-fir.sourceforge.net, 2005

All quoted internet resources were accessed and verified on Dec. 4th, 2007.

Free Ambisonic material can be downloaded from Etienne Deleflie's community repository at http://ambisonia.com.