Scaling up: making point-source multichannel content work for large listening areas

J. Nettingsmeier[1], P. Maier[2]

[1] *Verband deutscher Tonmeister e.V. Email: nettingsmeier@tonmeister.de*
2 concept-A acoustics GmbH, Email: peter.maier@concept-a.net

Abstract

Audio-related conferences face the challenge of presenting discrete multichannel content to large and demanding audiences of 100 or more listeners, with a high degree of fidelity. Such content is usually designed for small-area domestic or studio listening, assuming single point sources per channel, with well-defined level and arrival time relationships at the listening spot. These conditions are impossible to maintain over large areas, and suitable trade-offs must be identified. At the same time, acoustic deficiencies of typical large conference venues may severely distort the intended listening experience.

We describe the acoustic treatments, loudspeaker designs and signal processing methods used at the VDT International Convention (Tonmeistertagung, tmt) and the compromises made to scale up systems ranging from 2.0 to Auro 3D. These approaches are contrasted with methods used in cinemas, and with Ambisonic rendering.

Introduction

Conference venues are typically not suited to high-quality sound reproduction. Hence, room treatment is of fundamental importance, and the resulting acoustic sets the limits of what can be achieved. Within these limits, audio quality is largely constrained by physics and psychoacoustics. Their interplay must be understood, so that different aspects of fidelity can be weighed against one another, and optimal trade-offs found.

Quality criteria

The main perceptual factors to consider are tone colour, clarity, localisation, and balance.

Tone colour

We can assume any well-designed speaker to be nearly perfect for a near-field, on-axis listener. To control the tone colour for a distant listener in large reverberant venues is to control the diffuse field. This is fundamentally impossible without acoustic treatment of the room.

The speaker radiation pattern is another key factor. Most speakers will emit low frequencies omnidirectionally and only reach their specified dispersion angle at around 1 kHz, narrowing further as the frequency increases. Thus, low frequencies will dominate the diffuse field in the room. As we add more speakers, the cumulative effect is a muddy, boomy, bass-heavy sound despite perfectly linear on-axis response.

The seat dip effect (a narrow-band attenuation of sounds between 100 and 300 Hz with grazing incidence across rows of seats) does not mitigate this boominess, as it mainly attenuates early sound at the cost of clarity [1], and disperses rather than absorbs bass energy [2].

Somewhat counter-intuitively, wide-dispersion speakers may yield more musical results in a reverberant environment because of their more even diffuse field, trading clarity for tone colour.

Amidst a large audience, the heads and torsos of seat neighbours will create an acoustic shadow for shorter wavelengths, manifesting in a lack of direct sound in the high frequency band. Slightly elevating the speakers for uninterrupted sight lines from each listener to each tweeter improves spectral balance and clarity throughout the listening area. The resulting spurious height impression is less obtrusive than coloration from tweeter occlusion.

After aligning all speakers and correcting the room acoustics as much as possible, we must consider both the direct sound of each single speaker as well as the cumulative spectrum of all speakers during an integration window of the order of the reverberation time. Individual speaker equalisation can correct for speaker placement artefacts (usually a bass boost due to proximity to one or more boundary surfaces) and counter obnoxious room resonances which remain after acoustic treatment. Global equalisation can then trade near-field “punch” for a slimmer, more natural far-field impression by gently attenuating the bass range of all speakers.

Clarity

For musical purposes, clarity is defined as the ratio of sound energy in the first 80 ms versus the remaining energy until complete decay. Well-received concert halls have C_{80} ratios between -4 and 1 dB. [3]

Clarity may be secondary to tone colour for listening enjoyment, but insufficient clarity may lead to unconvincing demonstrations and lack of emotional impact. Clarity is paramount for movie dialogue intelligibility, or when the lecturer is reinforced over the same loudspeaker system. In those cases, the speech clarity C_{50} (with a reference window of 50 ms) might be a better metric.

Clarity can be improved by increasing the direct sound with more directional sound sources such as line arrays or horn systems, or by using delay lines. However, the most effective approach in the context of music reproduction is to reduce the reverberant field with acoustic absorbers.

Localisation and balance

Most recordings tacitly assume a single listener seated in the centre of the loudspeaker system. Under this condition, the two most relevant sensory cues for source localisation can be reproduced: interaural level difference (ILD), and interaural time difference (ITD). This leads to a perceptually correct rendition of a soundstage that spans the entire speaker base without appearing to originate from the loudspeakers, and provides natural and sharp localisation without listening fatigue.

However, ideal stereophonic reproduction is very fragile. ITD localisation operates in a time window of +/- 1.5 ms [4], which translates to about +/- 0.5 m of path length difference before the image collapses into the nearest speaker. Hence, correct ITD reproduction can only be expected on the median axis of the auditorium – even adjacent seats will suffer from significant directional distortion, and for places further to the sides, ITD localisation will break down completely.

ILD localisation is slightly more robust: in order for a sound to appear to come entirely from one speaker, a level difference of around 18 dB is necessary. Assuming perfect point sources in a free field, this translates to a path length ratio of 1:8. [5, 6]

Evidently, attempting precise ITD reproduction over a wide area is a futile exercise (Fig. 1). But even reasonably correct level ratios are difficult to obtain (Fig. 2). For off-centre listeners, the sound image will inevitably be pulled towards the nearest speaker.

Only very few listeners will experience the perfect equilateral stereo triangle and thus correct localisation in the absolute sense. In practice, sufficient relative localisation can be maintained as long as the image does not collapse into a single speaker (red areas in Fig. 2), phantom sources can be separated (i.e. the angle between speakers is still wide enough) and phantom images remain stable (the angle is not too wide).

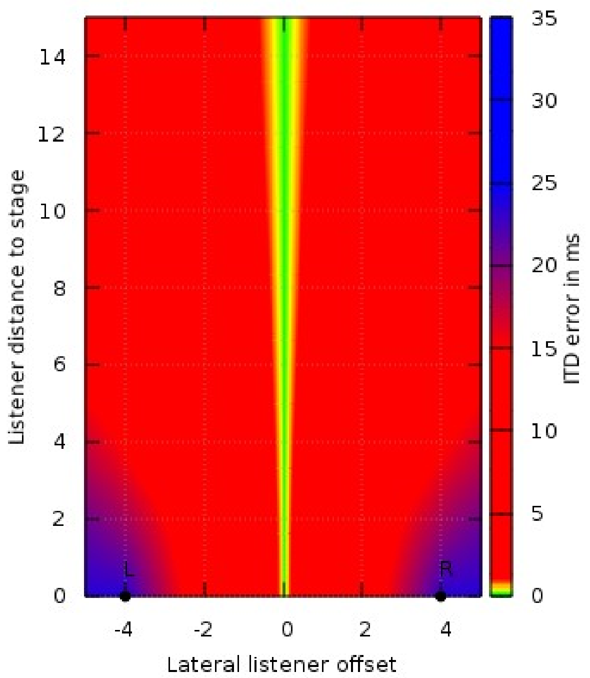

Fig. 1: ITD error for a 10 x 15 m venue. Ideal omni-directional loudspeakers spaced 8 m apart, at bottom. Green denotes areas with less than 50% localisation error. In the red areas, the nearest speaker dominates localisation. Blue areas may approach the echo threshold for signals panned to the opposite side.

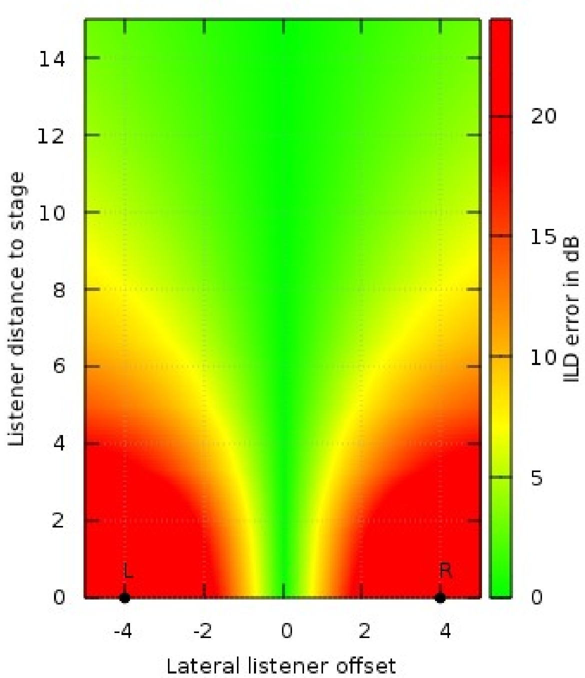

Fig. 2: ILD error for the same venue.

All these considerations pertain to phantom sources only, that is to say sounds which are played back over two speakers with a high degree of correlation. Uncorrelated signals are always perfectly localisable at the location of their speaker, but they will appear as singular sounds that do not blend into any continuous sound stage.

Room acoustics

The most fundamental acoustic requirement is to reduce reverberation, which must not be reduced to a single average decay time, but needs to be considered per frequency. Additionally, we strive for controlled reflection behaviour, homogeneous decay throughout the spectrum, and thus a perceptually neutral transfer function.

Room defects must be addressed individually. Flutter echoes between parallel surfaces are highly detrimental to music reproduction, and compromise speech intelligibility. Pronounced single reflections might impair localisation or colour the sound if early enough to cause comb filtering. Room modes lead to uneven decay, and their stationary nature may cause grave fluctuations in tone colour across the listening area.

In fixed installations, a combination of absorbers, diffusors, and specular reflection surfaces is used to control these issues. For temporary setups however, the focus will be on absorption, usually via textiles and foam material, for ease of installation, rigging and transport.

Absorbers should be designed to cover a sufficiently wide frequency range to avoid coloration of the decay. Good low-frequency attenuation is paramount, as it reduces ringing of the room modes. Narrow-band absorbers can be tuned to particularly troublesome resonances.

The addition of extra speakers (or whole layers of speakers) increases the number of acoustical transmission paths to be considered, and effectively rules out easily implemented two-speaker concepts such as the live-end/dead-end design.

With a larger acoustically relevant surface area, the required amount of absorbent material increases. Fortunately, this affects mostly the mid and high frequencies, which can be approached with textiles. Low-frequency attenuation usually affects the whole room rather than individual transmission paths, so the effort required is constant regardless of the number of speaker channels.

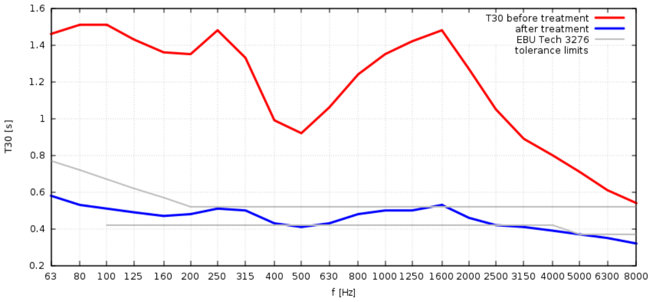

Fig. 3: The reverberation time of room R3 at tmt27 before and after treatment. The gray lines show the tolerance range extrapolated from EBU Tech 3276.

Acoustic treatment

The reverberation time plot of a room is a good starting point to design acoustical measures. The red line in Fig. 3 shows the measured reverberation time T_{30} in in the empty room R3 at Tonmeistertagung 2012. The measurements were obtained with an impulse source.

Different target values and tolerance ranges are available, such as DIN 15996, DIN 18041, the Dolby and THX requirements for cinema or the recommendations for control rooms EBU Tech 3276 and ITU-R BS 1116.

Since room R3 is mainly intended to reproduce professional audio productions, we choose the EBU recommendation.[7] It is limited in scope to smaller rooms, but since it defines the allowable reverberation as a function of the room volume, we can extrapolate:

$$\mathbf{T_m = 0.25s \cdot (V/V_0)^{(1/3)}} \quad [s] \qquad (1)$$

Here, T_m is the nominal reverberation time, V_0 is the reference volume for a listening room ($100m^3$), and V is the actual volume, in our case $684m^3$. The nominal reverberation time is used to compute a frequency-dependent tolerance band, which is shown in grey in Fig. 3.

Along the sides and rear of the room, three layers of black heavy stage cloth run from the floor up to a height of 3m, attached to pipes which are partly supported by tripods and partly by adjustable steel wires suspended from the ceiling.

The spacing between cloth and wall results in wide-band absorption characteristics for diffuse sound down to low-mid frequencies. In this particular case, the spacing is between 30 and 70 cm. The drapes cover all critical surfaces that might cause unwanted reflections or flutter echo.

Resonant wide-band absorbers at the frontal wall and along the side walls behind the curtains further increase absorption in the low-mids and are tailored to smoothen the frequency-dependent decay identified by the T_{30} measurement.

Edge absorbers along the vertical and partly along the horizontal room edges behind the curtains attenuate the low-frequency response of the room. To address individual critical reflections of doors, technical installations etc., mobile wideband absorber partitions are used.

The ceiling is left untreated for practical reasons. For future conferences, we will consider inflatable membrane absorbers. [8]

The blue curve in Fig. 3 shows the reverberation times after treatment.

Format-specific considerations

Stereo

As shown above, large listening areas exacerbate the lateral stability problems that have plagued two-channel stereo since its inception and have ultimately led to the design of three-speaker systems such as [9]. The only practical approach to this limitation is to make sure that both the presenter and the audience are aware of it, and to allow time for people to move to the median axis when ITD-related effects are to be demonstrated.

AB stereo recordings are particularly brittle as they contain no ILD cues at all. In the red areas of the ITD plot in Fig. 1, phantom sources can be expected to collapse into a single speaker.. Equivalence techniques such as ORTF provide enough ILD for good localisation over most of the green area of Fig. 2, as do coincident or pan-potted stereo recordings.

Sometimes, different stereo techniques are combined in a single production, and the outcome on a large system may be surprising even to the producer. It is generally helpful to make sure in advance that the presenters are aware of these physical facts, to avoid chasing ghosts during sound check.

That said, “good enough” stereo can be scaled up almost arbitrarily, using large line arrays or clusters capable of delivering near-constant sound pressure level over long distances. Commercial deployments usually favour a dual-mono approach, where each array covers one half of the audience, but this is not mandatory, and systems with sufficiently wide opening angles are available. However, due to their phase-coherence and dominant direct sound, overlapping line arrays may exhibit stronger phasing with correlated content than point sources.

Needless to say, the perceived tone colour of a large system will change dramatically along the length of the coverage area, but since it matches the visual impression of distance, the result is usually accepted as convincing and adequate. Large line sources may even cause listener irritation because their sound field is more precise and near-field-like at great distances than expected.

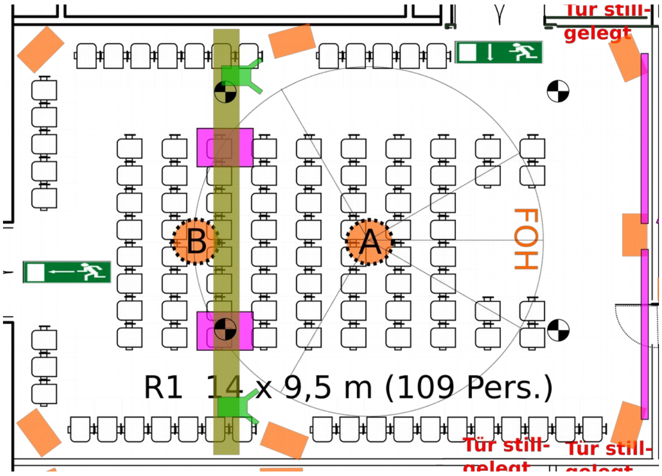

Fig. 4: 5.1 example from tmt27, 2014. This is a workshop room, so the FOH position is actually in front of the audience. The orange boxes are speakers.

5.1/7.1

With the addition of a centre speaker, the area of precise stereophonic reproduction is widened to the point where larger audiences begin to make sense. The width of the L-C and C-R stereo bases is halved compared to two-speaker stereo, and the usable area increases correspondingly. The centre speaker will provide stable central localisation even for listeners off to the sides, provided the material has low crosstalk.

The remaining challenge is to strike the right balance between front and surround speakers, to maintain stable frontal localisation while providing satisfactory envelopment at the same time.

Fig. 4 shows the approach we have taken for the VDT International Conventions 2012 and 2014:

- Speakers are placed as far away from the audience as practical, and close to boundary surfaces to provide low-frequency support.
- We define a golden seat ***A*** in the front half of the auditorium that has a slightly too wide LR angle (acceptable because of the increased stability due to C, and beneficial for the rear part of the room).
- L, C, R, and the surround speakers Ls1 and Rs1 on the side walls are calibrated for equal levels and arrival times at seat ***A***.

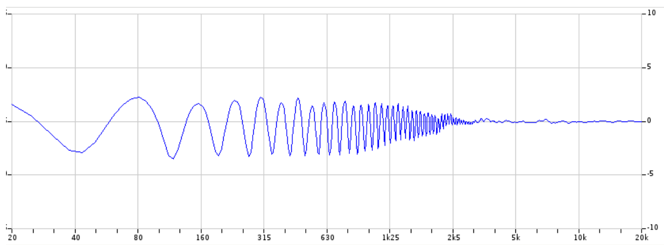

*Fig. 5: Theoretical comb filter resulting from the spill of the rear surrounds into the frontal listening zone around **A**.*

- We define the “golden seat among the cheap seats“, ***B***, in the rear part of the auditorium, and measure the average incident sound level and time of arrival from L, C, and R. Let's assume a level drop of -3.5 dB and an additional delay of 8.7 ms relative to ***A***.
- Additional surround speakers Ls2 and Rs2 in the rear corners are calibrated to said level and arrival time for good balance in ***B***.
- We consider the impact of Ls2 and Rs2 on the golden seat ***A***: the spill arrives approximately 13 ms later, at a level of -7dB. The resulting coloration only affects mid to low frequencies and is mild in comparison to other room effects (Fig. 5). Since it arrives late, it will be fused into a single auditory event localised at Ls1 and Rs1. [10] The goal is to control the spill level and achieve a time-of-arrival difference that is long enough not to cause audible comb filtering, and shorter than the echo threshold.
- Depending on the dispersion pattern of Ls1 and Rs1, the surround image for the cheaper seats around ***B*** may be drawn to the front. In that case, the delay for Ls2 and Rs2 can be reduced while checking the frontal seating area for coloration.

This approach ensures that participants willing to arrive early can secure seats with near-perfect reproduction, while a larger part of the audience is getting at least a decent surround experience without major impairments.

If required, we provide an additional 7.1 preset with different delays and gains, where the rear surrounds are driven discretely, and the sweet spot is shifted towards the rear. We advise the presenter and audience that only the central seats will have useful coverage, and that the image may collapse into the nearest speaker for peripheral seats.

In cinema, surround speakers are multiplied more aggressively, with each pair only covering a few rows. The surrounds have a base delay of 10ms in addition to any distance compensation and are either wired in parallel to save cost, or individually amplified. In the latter case, they can be timed such that the speakers in the back hit the golden seat first (which in cinema is typically located 2/3 to the back of the room), to anchor the image to the rear. The other speakers then follow back-to-front. This trades uniform level, stable rear localisation and decorrelation for blurred transients and some degree of comb filtering in the surrounds. It might give inconsistent results if the material contains important percussive content rather than just ambience, or if LCR and surround channels are correlated. The latter is true for most music recordings, or when sources are panned between front and rear. [11]

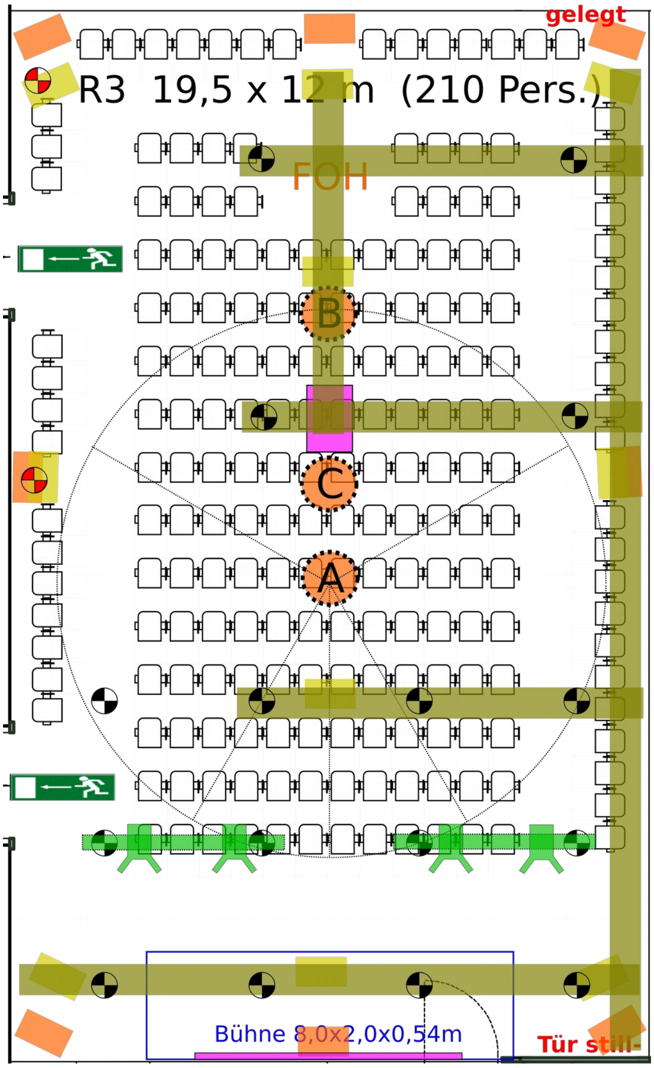

Fig. 6: Auro 3D 13.1 example. Orange squares are floor speakers, pale yellow ones are height speakers, attached to trusses or individual rigging points.

Auro 3D

For Auro 3D, we are taking the same approach as for 5.1. A frontal sweet area is extended towards the rear by an additional group of floor and height surround speakers. The optional ceiling speaker can be doubled as well and the rear one attenuated and delayed accordingly (Figure 5).

As an additional constraint at the Tonmeistertagung, the same setup doubles as a Higher-order Ambisonic playback system. To improve coverage for Ambisonics, the first group of surround speakers is not angled towards ***A***, which causes significant spill into the rear seats and pulls the surround image to the front. Reducing the delays of the second group of surround speakers ensures a stable rear image.

We provide Auro 13.1 capability in the same way we provide 7.1, via discrete access to the rear speakers, with similar disadvantages.

Fig. 7: The Auro 3D/Ambisonics room at tmt28, featuring two rings of eight and two ceiling speakers, two frontal subwoofers and one at the rear. All loudspeaker management tasks are performed by the console visible at the bottom. Acoustic curtains on three sides, with additional absorbers behind them.

All hemispheric with-height systems will pull horizontal sources slightly upwards in the presence of crosstalk into the upper channels. Recordings using omnidirectional microphones for the top layer are particularly affected, and even low-crosstalk microphone arrays are not completely immune.

In most manufacturers' speaker installation guides, we find that height speakers should be placed directly above their floor counterparts without compensating for the additional distance (thus arriving late), and angled to aim slightly above the head of a central listener (for a little bit of treble attenuation), likely to address the pull-up effect. The same techniques can be used in large systems, but their effect will be very subtle due to the larger speaker distances.

We conjecture that two other factors can mitigate these problems and improve listening enjoyment:

- A wider dispersion angle in the height speakers seems to be desirable. We found this out accidentally when we had to combine horn-loaded systems on the floor with flown dome tweeter systems without waveguide, due to limited supplies.
- A more “live” upper half of the room may be beneficial. Our room treatment ended 3 m above ground, but rather than causing issues, the extra reflectivity around the top speakers seemed to improve envelopment and reduce pull-up.

Higher-order Ambisonics

For Ambisonic rendering, we augment the Auro 3D setup to two rings of eight speakers (adding a rear floor and height centre and rear surround height speakers) and use the two ceiling speakers individually. We define a third reference point ***C*** in the centre of the room, to which we calibrate all levels and delay times.

The ITU 5.0 base layout with its very uneven angular resolution is not optimal for Ambisonics, but a separate

Ambisonic system may not always be feasible and/or economical. The additional rear speakers mitigate the problem, but large holes at the sides remain. For low-order content, it might be advisable to skip the front and rear centres to avoid excessive crosstalk.

Since a phase-coherent velocity decode (r_V) is only necessary in the low frequency range up to a few hundred Hertz, group delay is not critical for the largest part of the spectrum, where the energy sum of the speakers is relevant (r_E). Thus individual equalisation can be used in the upper band to match the sound characteristics of the speakers due to design differences, alignment (which might be optimised for other content and other sweetspots), proximity to boundary surfaces, or distance. Slight treble attenuation can prevent closer speakers from standing out too much. [12]

Correct sound field reconstruction can only happen in areas covered by all speakers. In most spaces, the side and top speakers define the limits of the listening area due to their proximity. If the source material is of sufficiently high order, this is the dominant limiting factor.

Ambisonic systems have high crosstalk between speakers, which can lead to audible coloration and phasing issues, and impairs transient response outside the sweet spot. With these base limitations in mind, it has been found to scale rather well [13, 14] while maintaining good localisation and envelopment, especially if the level can be held constant across the listening area using horn systems [15] or line arrays. [16]

Bass management

The higher the channel count, the more reason to use smaller speakers with limited low-frequency response, which calls for bass management to ensure satisfactory reproduction. The following guidelines have produced good results:

- Due to the distance between mid-hi speakers and subwoofers, the time alignment in the transition band will be wrong for almost all seats. Gentle crossover slopes would cause very uneven response across the listening area, with ripples over several octaves. Steep crossovers (such as a 24dB/oct Linkwitz-Riley) are preferrable, to keep the uncontrollable transition band as narrow as possible.
- For surround systems, four subwoofers in the corners of the room are highly desirable, to provide at least approximately correct directional information in the bass. Mid-hi speaker feeds can then be routed to the nearest one.
- Obviously, but easily overlooked, mid-hi speaker feeds must be tapped prior to their output delay, and be fed into the delay stage of the subwoofer busses for correct time alignment.
- It is possible to create phantom woofers for the front and rear centre and side speakers using a $\sin^2/\cos^2$ panning law for constant power, but no miracles should be expected.

We have found it convenient to implement the bass management in the mixer, but it requires a state-of-the-art product with tunable pass filters and flexible matrix routing.

Alternatively, a commercial DSP platform or even a general purpose computer with copious amounts of I/O and low latency could be used.

Seating arrangement

Many venues call for a middle aisle in their default floor plan. We strive to change this to a central block with side aisles to maximise the seat density in the sweet area. Since this will affect house fire safety and evacuation plans, alternative seating arrangements may need to be developed well in advance in cooperation with local authorities.

Standing-room audience should be avoided because of loudspeaker obstruction, which would impair even the optimally covered areas. Instead, we set up auxiliary chairs along the walls, facing inward, to allow latecomers to at least hear the presentation, while making the sonic limitations of these seats obvious (Fig. 4, Fig. 6). This appears to contribute to participant satisfaction.

Conclusion

For audio-related conferences, stick to the point-source paradigm as much as possible to maintain compatibility with classical surround recordings. Signal processing is only used after acoustic measures have been implemented. We try to deliver near-ideal reproduction in the frontal part of the listening area, and to degrade gracefully in the rear seats in terms of localisation and sound pressure level. Unlike laypersons, who might expect a uniform (if compromised) experience throughout the listening area, audio professionals tend to judge the system by its peak performance in the sweet spot and intuitively understand its limitations. Obvious “bad seats” create overflow capacity yet help steer the expectations of the listeners seated there.

So far, this approach has been met with a high degree of acceptance from presenters, audience, and hardware sponsors alike.

[1] William J. Davies, Trevor John Cox, and Yiu Wai Lam: *Subjective Perception of Seat Dip Attenuation*, Acta Acoustica united with Acoustica, 1996

[2] Tapio Lokki, Alex Southern, and Lauri Savioja: *Studies on seat dip effect with 3D FDTD modeling. Aalborg, 2011*

[3] F. Alton Everest, and Ken C. Pohlmann: *Master Handbook of Acoustics*. 5th ed, McGraw-Hill, 2009, p. 386

[4] Jens Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, 1997, p.144

[5] Eberhard Sengpiel, G*leichungen für die Pegeldifferenz- und Laufzeitdifferenz-Lokalisationskurve*. http://www.sengpielaudio.com/GleichungenDLundDt.pdf, 1995

[6] Helmut Wittek, *Lokalisationskurven online – Der Image Assistant.* http://hauptmikrofon.de/HW/IMA-Doku.pdf, revised 2001

[7] EBU Tech. 3276: *Listening conditions for the assessment of sound programme material: monophonic and two–channel stereophonic.* 2nd ed. 2008, https://tech.ebu.ch/docs/tech/tech3276.pdf

[8] http://www.gerriets.com/media/wysiwyg/downloads/technic/4-1-aQflex-akustiksystem_3410.pdf

[9] International Telecommunications Union, *Recommendation ITU-R BS.775-3, Multichannel stereophonic sound system with and without accompanying picture*, revised 2012

[10] Helmut Haas, *The Influence of a Single Echo on the Audibility of Speech.* JAES, 20:2, pp. 146–159, 1972, after [4]

[11] Howard Massey (ed.):*The Recording Academy's Producers & Engineers Wing Recommendations For Surround Sound Production.* The National Academy of Recording Arts & Sciences 2004, section 2-4 to 2-8. http://www2.grammy.com/pdfs/recording_academy/producers_and_engineers/5_1_rec.pdf

[12] Jörn Nettingsmeier and David Dohrmann, *Preliminary studies on large-scale higher-order Ambisonic sound reinforcement systems.* Ambisonics Symposium, Lexington KY, 2011. http://stackingdwarves.net/public_stuff/linux_audio/ambisonic_symposium_2011/AmbiSym2011-Nettingsmeier-Dohrmann_Large-scale_HOA_Systems.pdf

[13] Dave Malham, *Experience with Large Area 3-D Ambisonic Sound Systems*, in: Proceedings of the Institute of Acoustics Autumn Conference on Reproduced Sound 8. Windermere 1992 pp.209-216. http://www.dmalham.freeserve.co.uk/ioapaper1.pdf

[14] *A dance event at the University of York*, York 1996, http://www.york.ac.uk/inst/mustech/3d_audio/ambix/ambisoni.htm

[15] Bruce Wiggins, *WigWare @ Glastonbury*, Blog entry, July 5 2010. http://www.brucewiggins.co.uk/?p=33#content

[16] Nettingsmeier and Dohrmann, l.c.