REAL TIME WALKTHROUGH AURALIZATION - THE FIRST YEAR

B.-I. Dalenbäck CATT, Mariagatan 16A, Gothenburg, Sweden
M. Strömberg Valeo Graphics, Seglaregatan 10, Sweden

1 INTRODUCTION

Various limited forms of auralizations with real time movement and/or head rotation have been reported during the last decade. Examples of approaches have been:

- direct sound only (according to the definition[1] this can not really be called auralization);
- direct sound followed by an Infinite Impulse Response (IIR) reverb (similarly, no room model is used so not quite auralization either);
- direct sound and early reflections followed by an IIR reverb[2];
- direct sound and early reflections in a simplified room followed by a static pre-calculated Finite Impulse Response (FIR) reverberation tail from the detailed model, based on B-format[3];
- for fixed positions but with head-tracking, an early example was based on an 80 ms long early part with head-tracked binaural FIRs followed by a static reverberation tail FIR[4];
- for fixed positions but with head-tracking, a more recent example based on full auralized or measured binaural FIRs[5].

The technique presented here, as implemented in the *CATT-Walker*™ module of *CATT-Acoustic*™ [6] is based on full-length B-format FIRs and allows for free movement as well as head rotations. This paper discusses the technique, its properties, limitations, potential pitfalls, current and future possibilities.

2 TECHNOLOGY

2.1 Basics

The technique used is in principle simple and uses parts that have been around for quite some time:

- full length FIRs for auralization are created in B-format (W,X,Y,Z), available in *CATT-Acoustic* since more than a decade[3]. The use of B-format allows for arbitrary sound field rotations requiring only one B-format FIR set for each position. In contrast, a direct binaural simulation would require a large set of full-length binaural impulse responses for every single receiver position making it practically impossible with a walkthrough;
- continuous real time interpolation and convolution of full length FIRs. Low latency convolution is licensed from *Lake Technology* [7], where the uneven FIR partitioning enables a lower latency and less CPU than fixed sized FIR partitions;

- a binaural down-mix of an ambisonic decode. As discussed below, a loudspeaker output would be simpler and is as well possible but a binaural decode was chosen since most acoustical consultants do not have a listening room set up for multi-channel replay while a binaural replay, using the same Head-Related Transfer Functions (HRTFs) and headphone equalization as is used with their normal static auralization, is straightforward to use;

- graphics and audio processing integration using *OpenGL*[8].

The complete process is shown schematically in Fig. 1 while the real time part is shown in Fig. 2. The process creates a seamless binaural simulation while navigating through the virtual room.

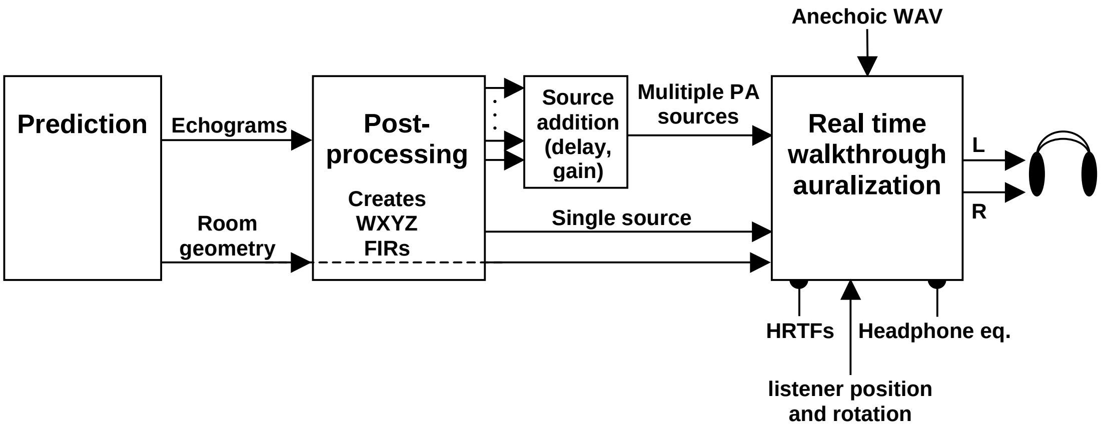

Fig. 1 Schematic flowchart describing the complete processing

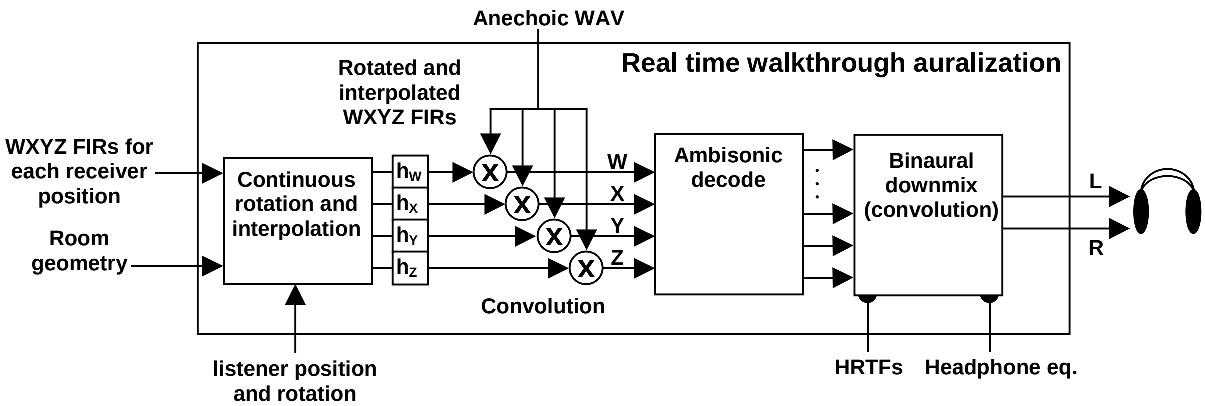

Fig. 2 Schematic flowchart describing the real time processing part

Sample screenshots from a church simulation can be seen in Fig. 3 and 4 where in Fig. 4 the receiver grid as well as an interpolated B-format FIR are shown (FIR drawing simplified for efficiency).

2.2 Current Options

Since this paper describes a commercial product, the feature discussions have been limited to those of general interest and related to the possible use of standard PCs. Along these lines, the basic currently implemented acoustically interesting options for the walkthrough are:

- multiple source simulation. Assuming that the sound input is common for all sources, such as in a PA system where a voice is amplified with a multi-speaker system, any number of sources can be simulated at no extra CPU cost since the impulse responses for all sources can simply be added before processing (taking relative delays and gains into account);

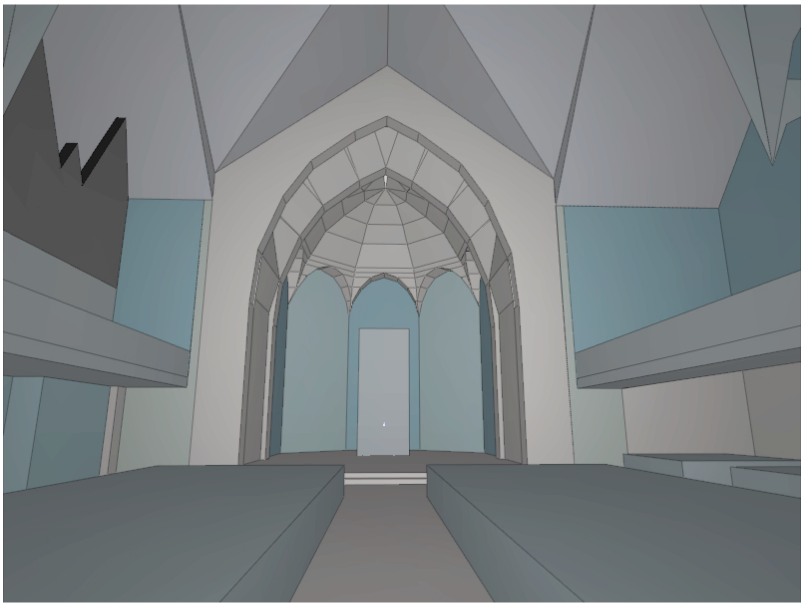

Fig. 3 Typical screenshot during walkthrough in a church

- choice of HRTFs for the binaural down-mix;
- choice of headphone equalization for the binaural down-mix;
- variable walking speed;
- optional TCP/IP control via the *Walker Steer* API. This makes it possible to have the graphic walkthrough run on another system (e.g. a VR system with full detailed graphics) where only the acoustics rendering is made by *CATT-Walker*.

Options related to the use on ordinary, even several years old, PCs are:

- processing block size and queue can be traded off against CPU use, especially on slower PCs a tradeoff with a higher latency can be made. Since the typical application is not VR, a very low latency is not crucial;
- the simulation can optionally work only horizontally without height (the Z component) thus consuming less CPU in both the convolution stage and the binaural down-mix stage.

2.3 Future Options

A number of additions and improvements are prepared and possible but some of them may not be of primary interest for pure room acoustics use.

- detailed models with textured graphics can be used as imported from programs such as *3ds max* [9], the only major requirement is that the acoustic model and the detailed graphic model use, or can be transformed to use, the same coordinate system. A potential use is for presentations together with an architect;
- head-tracking at no extra CPU cost. Due to the use of B-format, additional rotations require only adding an head angle value to the already performed normal rotations. However, since the typical use is in front of a PC screen, where there naturally are small head movements, head-tracking does not give any big additional advantages for normal use;

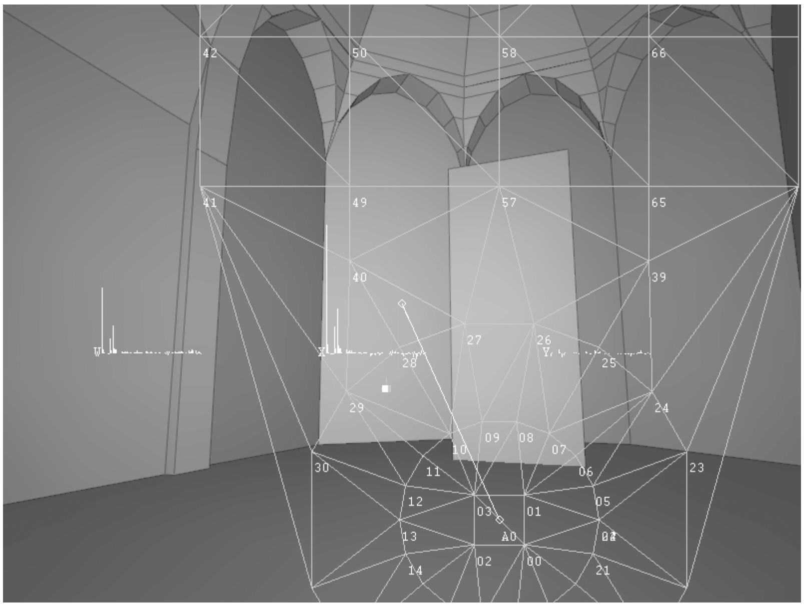

Fig. 4 Screenshot with additional display of receiver grid and impulse responses

- multiple independent sources. The CPU on a modern PC is powerful enough to run several independent sources (i.e. sources with different anechoic inputs) together but it will of course consume CPU in proportion to the number of sources;

- Doppler effects. Since the simulation is based on room FIRs, the frequency shift for individual reflections, that may be in any direction, cannot be handled. Because of this a less detailed technique have to be used. However, Doppler is of most interest in VR since for pure room acoustics walkthroughs the de-tuning of instrument tones can be very disturbing;

- direct B-format output for external decoding to any loudspeaker array. This option lowers the CPU demand as compared to a binaural output since the binaural down-mix of an ambisonic decode need not be performed;

- ambisonic output for direct loudspeaker replay, including height. Also this option lowers the CPU demand since the binaural down-mix is not required;

- use of measured instead of predicted B-format FIRs. This is a straightforward option where the B-format FIRs can be measured by a *Soundfield* microphone[10] or equivalent;

- use of 2nd order B-format (five more FIRs/channels: R,S,T,U,V). This will require more than twice the CPU and can give a better localization and a wider sweetspot but mainly for loudspeaker replay. It is questionable if it is worth the extra CPU for a binaural down-mix since the current 1st order decode already gives a good localization (note that with a binaural down-mix, the listener sits exactly in a virtual sweetspot so the small sweetspot problem with 1st order ambisonics does not affect as much).

3 PROPERTIES, LIMITATIONS AND PITFALLS

An interesting property of the real time part of the process is that the audio processing is totally independent of the room complexity but depends only on the length of the FIRs. A walkthrough auralization of a simple shoebox with 2 s reverberation time will consume exactly the same CPU as a walkthrough in a detailed concert hall with 2 s reverberation time. The complexity of the model affects only the echogram prediction stage.

The prediction and post-processing methods are exactly the same as for static auralization and the FIRs will therefore have the same properties and be subject to the same limitations, i.e. as imposed by geometrical acoustics etc., but the additional use of interpolation presents some potential pitfalls:

- If a pillar or a wall is between receiver positions, interpolation between FIRs on both sides of the wall may happen. This can be avoided but the intended application is not room mazes and similar, that are anyway very difficult to predict and auralize well, but rather typical mainly open-plan rooms like auditoria, concert halls, churches, operas, classrooms etc.;

- Interpolation between positions means that the user has to use a high receiver density where the acoustics, and/or the direction to the source, is expected to vary rapidly with movement and/or head rotation. Typical high density areas are close to and around the source, where concentric rings of receivers are best used, but also positions around door openings and close to objects that may give significant reflections. The pitfall is that the interpolation can sound realistic and convincing *even* if the receivers are placed way to sparse but this can also be used for quick initial tests require less calculation time. Fig. 5 gives an example of typical receiver placement for a church walkthrough.

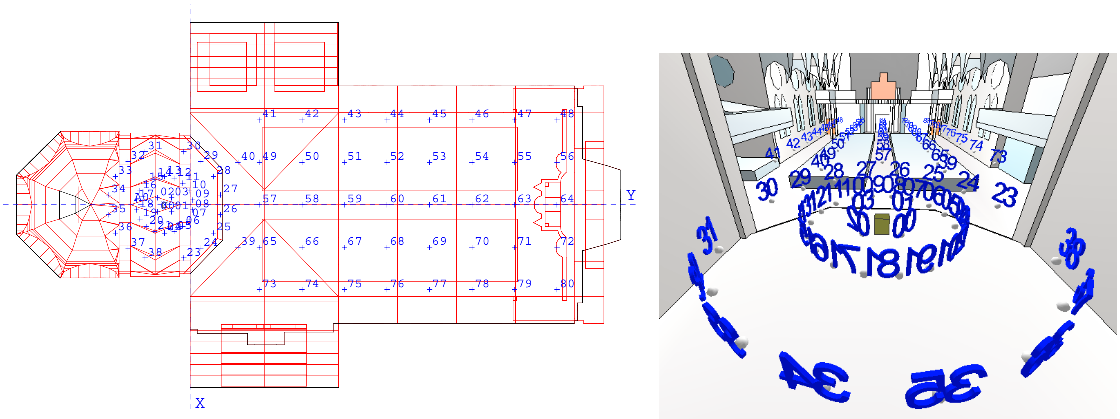

Fig. 5 Typical receiver placements for a church walkthrough

4 TYPICAL AND POTENTIAL USE

As compared to normal static auralization, the additional effort to use walkthrough auralization is only to use more receivers and to take some care about their placement and density as discussed in section 3. The higher number of receivers leads to slightly longer prediction and post-processing times but in relation to the possibility of doing a walkthrough it is marginal and can also run unattended. As an example with an “everyday” typical prediction case, such as a room with a 1 s reverberation time using 87 receivers, see Fig. 6, the prediction time is 6 min and the post-processing time is 32 min and from then on free real time movement is possible (calculation times from a three-year old 1.8 GHz *Mobile Intel Pentium 4* laptop). For the same laptop, the pre-processing for the 3.6 s reverberation time church is 39 min prediction, due to the more complex geometry and longer echograms, and 83 min post-processing due to the longer FIRs. The FIR data-file for the 1 s case occupies ca. 30 MB (11 MB compressed) while the 3.6 s case occupies ca. 100 MB (40 MB compressed). From these sizes it can be understood that the PC used for the walkthrough typically has to have 512 MB memory, which is rather standard.

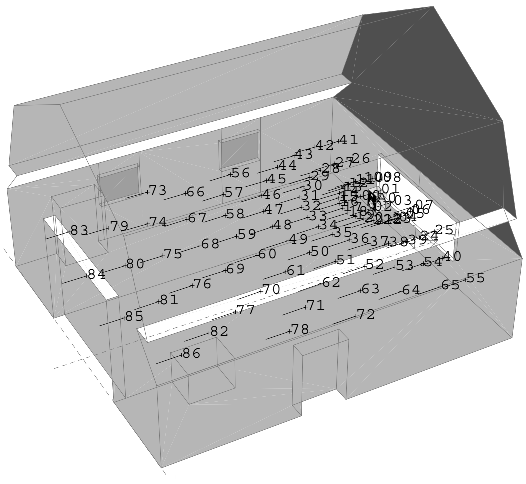

Fig. 6 A basic “everyday” room acoustics prediction case with 87 receivers placed for real time walkthrough

If the new walkthrough technique will be mostly used internally by consultants, or together with project presentations, remains to be seen. For presentations, it may instead be preferable to create a detailed rendered video walkthrough in e.g. *3ds max* and synchronise auralized soundtracks created via the off-line Walkthrough convolver, a technique available since half a decade.

Further exciting uses explore the possibility of control via TCP/IP. Examples of such uses are with a large-screen simulator for the EC “POEMS” project at *Chalmers University*, Gothenburg[11] and with an immersive virtual environment for the EC project “Wayfinding” at *LIMSI* in Paris[12].

5 CONCLUSION

A technique for real time walkthrough auralization has been described, the technique is based on B-format impulse responses, is in itself general and can as well be based on measured responses. Future improvements of prediction and auralization methods will directly carry over to the walkthrough auralization since no special shortcuts need to be made for the real time option.

6 REFERENCES

1. M. Kleiner, B.-I. Dalenbäck and P. Svensson, Auralization - an Overview, JAES 41(11), 861-875, 1993
2. J. Huopaniemi, L. Savioja and T. Takala, DIVA virtual audio reality system. Proc. Int. Conf. Auditory Display (ICAD'96), Palo Alto, California, Nov. 8-11, 1996, pp. 111-116.
3. A. Reilly, D. McGrath and B.-I. Dalenbäck, Using Auralization for Creating Animated 3-D Sound Fields Across Multiple Loudspeakers, preprint 4127 99^{th} AES Convention, 1995
4. B.-I. Dalenbäck and D. McGrath, Narrowing the Gap Between Virtual Reality and Auralization, Proc. 15^{th} ICA, 429-432 (1995)
5. C. Moldrzyk, Design Tool Auralisation, DAGA 2004
6. www.catt.se
7. www.lake.com.au
8. www.opengl.org
9. www.autodesk.com
10. www.soundfield.com
11. www.ta.chalmers.se
12. www.limsi.fr/Scientifique/ps/thmsonesp/SonEspace