AMBISONIC REPLICATION OF CONCERT HALL ACOUSTICS FOR SOLO MUSICIANS WITHIN A DIGITAL AUDIO WORKSTATION: INITIAL EVALUATION

B Watson School of Computing, University of the West of Scotland, Paisley
DF Clark School of Computing, University of the West of Scotland, Paisley

ABSTRACT

This paper assesses the reproduction accuracy of a system that replicates the acoustic conditions experienced by a musician in a specific auditorium. Spatial impulse responses and support ratio parameters are measured at the venue, and subsequent real-time convolution is utilised to recreate the acoustics in a rehearsal setting. Ambisonic reproduction is via four loudspeakers of a 5.1 array, with digital audio workstation software and associated plug-ins performing signal processing, encoding and decoding.

The work differs from related research by offering an integrated approach, relevant to the domestic listening environment. Balance of the direct and diffuse sounds within the simulation is implemented here, whilst it is left unexplained in other work. A range of listening tests is presented.

1 INTRODUCTION

It is well recognised that solo musicians interact closely with the specific acoustic conditions of the venues in which they perform. Studies into this area reveal that experienced soloists react to their environment via the careful modification of fundamental performance aspects including breathing, articulation, timbre, dynamic range and even enacting mechanical changes to the instrument (e.g. string and reed selection). The acoustic conditions are perceived to form an 'extension' of the instrument being played and thus, constitute an integral part of the musical performance.[1,2]

Despite the importance of the prevailing acoustic conditions, financial and logistic constraints impose limits on the amount of time available for practice within a concert auditorium. Whilst the more experienced performer will be adept in reacting to relatively unfamiliar acoustic conditions, this may not be the case for less experienced performers. As a result, the overall quality of the musical performance may be compromised. To overcome this problem, this paper describes and assesses an integrated system providing replication of acoustic conditions for specific auditoriums within a digital audio workstation (DAW) environment.

The ambisonic system for surround reproduction was developed in the 1970's by M. Gerzon et al, offering a sophisticated and adaptable method of recreating the sound field around the listener.[3,4] Unfortunately, due to high hardware costs at that time commercial success was not realised. Advances in digital technology have now reduced the obstacles of cost and complexity, and as a growing number of audio engineers and musicians have access to a digital audio workstation, multichannel audio interfaces and loudspeaker array, then so the practicality of 'domesticating' such technology increases. This study aims to address the historical difficulties associated with ambisonic reproduction[5] and demonstrate that it can be put to creative use within a domestic setting.

Ambisonic reproduction is known to yield a relatively large listening 'sweet spot' when compared with discrete channel systems such as Dolby 5.1, and so is more forgiving in regard to the expressive movements of the performer that constitute an important part of the performance.

Steering away from pin-point accuracy in panning laws ensures that acoustic reflections are shared in appropriate proportions between all loudspeakers in the array and not just two, as is the case with the pair-wise panning employed by discrete channel systems. Therefore, the sound field is likely to remain stable even if the player moves their head and offers a great deal of immersion within the sonic environment.[6]

The work presented compliments previous studies into the development of acoustic environment simulations[7,8,9] and presents new approaches for the implementation of such a system in respect of the ease of impulse response measurement, calibration of the reproduction array and integration into a DAW environment. Whilst the systems cited above tend to use highly specialised convolution processors and larger loudspeaker arrays, this investigation assesses the viability of a more straightforward horizontal ambisonic reproduction employing consumer level equipment. The ability to switch between the perspective of the player and that of the audience, and the careful level balancing of the direct and reverberant sounds within the simulation are fundamental features implemented here that are not fully realised or discussed in existing research.

The papers cited above often assume that the typical musician would have an anechoic room at their disposal and moreover, that this could act as a suitable space for rehearsal. This work places the musician in a comfortable domestic rehearsal setting, dealing with issues regarding room resonances and 'room in room' acoustics from the outset.[10] Calibration of the system to adapt to different listening environments ensures that it is more readily suited to use in the rehearsal space.

2 SYSTEM SPECIFICATION

The system involves two principal stages, namely the measurement of spatial impulse responses in the venue and the subsequent real-time replication of the acoustics within a rehearsal environment.

2.1 Impulse response collection

An A-format microphone[11] is used to record impulse responses, and features capsules corresponding to the directions of LEFT-FRONT-UP, RIGHT-FRONT-DOWN, LEFT-BACK-DOWN and RIGHT-BACK-UP respectively. For stage measurements the microphone is positioned to represent the player’s head, 1.65m in height and 1.5m from the edge of the stage. For auditorium aspects it is placed 3m away, facing centre stage and at ear height of the seated listener (1.2m).

The room excitation signal is played over a Genelec 8020B active loudspeaker facing directly out into the auditorium at a height of 1.35m, representing the player’s instrument. The resultant impulse responses are to be used for instruments with similar directivity patterns, e.g. recorder, trumpet and voice. Other instruments would necessitate an alternative loudspeaker angle/height due to differing directivity patterns. Whilst this approach is at odds with the practices outlined in the relevant ISO standards document,[12] it does ensure a more consistent approach to room excitation since the instruments studied are broadly directional. This should prove more realistic than exciting the room equally in all directions with an omnidirectional source. Similarly, the use of a multi-directional receiver is more in keeping with the hearing capabilities of the musician.

Whilst it is anticipated that an application for the collection of impulse responses will be developed by the research group, measurements have thus far been carried out using the Apple Impulse Response Utility.[13] In the software, a four channel configuration is selected. A ten-second sine sweep is used as an excitation signal, and once deconvolved, each impulse response can be unpackaged from the proprietary project file and loaded into the convolution within the simulation.

The sine-sweep method is chosen for its superior performance in a quiet, unoccupied setting as it does not require calibration or suffer from the distortion peaks that can hinder alternative techniques.[14,15]

2.2 The real-time simulation

The real-time simulation employs the A-format microphone as a receiver, placed near to the head of the performer. The four channels are passed through pre-amplifiers on the audio interface with matched gain and routed to individual tracks within the inexpensive DAW application, Reaper.[16] A flow chart representing signal flow and processing for the system is given in Figure 1.

The four channels are then convolved with corresponding A-format impulse responses from the venue using the 'ReaVerb' Reaper plug-in and routed onwards for ambisonic processing.

Conversion from A- to B-format and ambisonic decoding is handled by freeware plug-ins from Daniel Courville[17] and David McGriffy,[18] although these will be succeeded by plug-ins developed by the research group. The output of the decoder is then routed to four loudspeakers of an ITU 5.1 layout, with the azimuth for each loudspeaker on the decoder adjusted to reflect the irregularity of the array.

Additional processing plug-ins will be devised for calibration of loudspeaker levels, room equalisation, feedback prevention and perhaps most importantly, the balance of direct and reverberant sounds in the real-time simulation by means of support ratio analysis.[1]

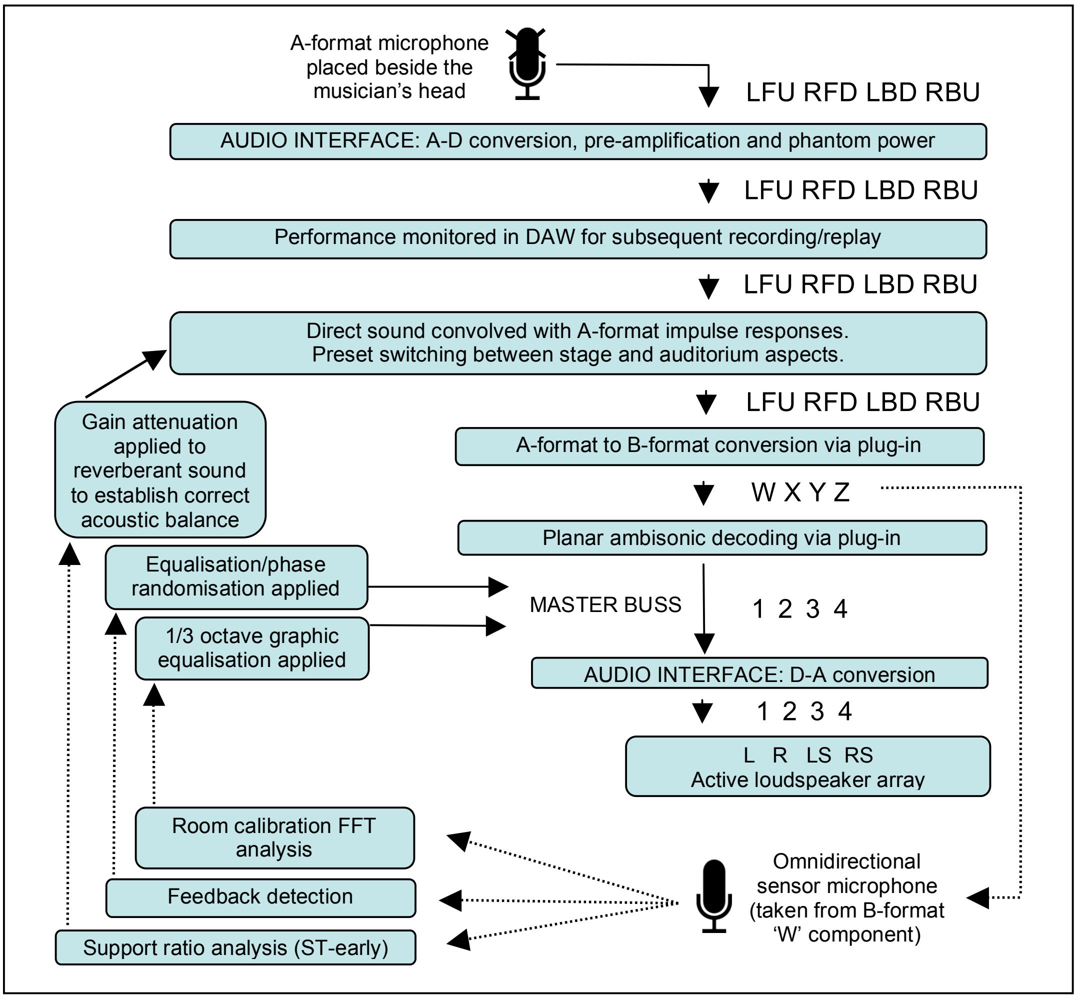

Figure 1: Flow chart outlining the real-time simulation system in a fundamental state.

3 OBJECTIVE TESTING

An early objective experiment concentrated on frequency response, replication of directional characteristics and reverberation time. This assessed the reproduction accuracy exhibited by the real-time system without the presence of room equalisation or feedback prevention.

In preparation, A-format impulse responses were recorded at three concert venues in Paisley, Scotland. Details of each venue are presented in Figure 2. Employing the techniques outlined above, two positions were measured in each venue for two instruments with differing directivity characteristics; violin and recorder.

Venue	Approx. Seated Capacity	Reverb Time at Position 2 (RT60)	Position 1	Position 2
The Brough Hall	200	1.5s	Stage	3m from stage
Thomas Coats Church	600	2.9s	Choir stalls	3m from choir stalls
Thomas Coats Hall	300	3.7s	Stage	3m from stage

Figure 2: Details of the three venues studied

The experiment was conducted in a listening room (measuring 6.5m x 7.5m x 2.8m) featuring acoustic treatment for limited high frequency damping; thus approximating a domestic setting. Genelec 8020B loudspeakers were set at ear height in an ITU 5.1 layout at a distance of 2m from the listening position (with only the L, R, LS and RS loudspeakers being utilised in the simulation). Loudspeaker levels were calibrated using pink noise and an omnidirectional microphone.

A 10-second sine sweep was fed in equal proportions to the four channels within the simulation and then each was convolved with the corresponding impulse response measured at the venue. Following conversion to B-format, the stream was decoded and output to loudspeakers.

The sound field microphone was placed at ear height at the listening position, and the output of the simulation recorded as an A-format stream. Finally, the sine sweep excitation signal was deconvolved from each of these recordings to reveal impulse responses that could be analysed and compared to original responses from the venue.

4 SUBJECTIVE TESTING

Subjective listening tests were performed using the same listening room conditions as the quantitative experiments detailed above. Ten expert subjects were selected for their experience in sound recording or performing acoustic music in concert venues, made up of department staff and honours year/postgraduate Music Technology students. Each was surveyed prior to the test to determine their familiarity with acoustic music as well as assessing exposure to loud amplified music which might have an impact on the subject’s perception during the testing phase.

Subjects were presented with pairs of extracts, with the first in each pair being an ambisonic recording of an instrument within the genuine acoustic. The second in each pair was a simulated performance consisting of a 'dry' version of the instrumental extract recorded in the listening room, convolved with the appropriate impulse responses from the venue. Care was taken to match dry A-format signals with the most appropriate responses taken from the venue, i.e. when replicating the auditorium aspect the right and left channel impulse responses were reversed since the dry extract was recorded using the 'stage' microphone position. The wet/dry signal balance at the output of the convolution plug-in was adjusted manually to best match the characteristics of the original recording. However, for subsequent testing a new plug-in will be used to analyse acoustic support and establish the correct balance during system calibration.

Two instruments were presented, corresponding with the specific violin and recorder impulse responses collected for objective experiments. For each venue, both stage and auditorium aspects were recorded and replicated. Subjects compared extracts and graded the simulated acoustics in respect of three parameters; reverberation decay time, timbre and immersion. The grading system is outlined in Figure 3, preceding a description of the instrumental extracts in Figure 4. In addition to numerical values, subjects were encouraged to add written comments to quantify their experiences.

5	Imperceptible – the reverberation qualities of the simulation extract match the original
4	Marginally perceptible – the simulation extract is very close to the original recording
3	Clearly perceived differences, but the simulation shares fundamental features with the original
2	Extremely perceptible – the simulation has limited similarities with the original
1	Substantial differences – the simulation extract bears no resemblance to the original recording

Figure 3: The numbering system used by subjects to grade differences between paired extracts

Instrument	Piece	Articulation & dynamic range
Violin	*Irish Reel* (Traditional)	Sustained/legato; narrow dynamic range.
Alto recorder	*Green of the Irishwoman* (David Gordon)	Mixture of staccato and legato articulations; wider dynamic range.

Figure 4: Comparison of the two instrumental extracts used for listening tests

5 RESULTS

5.1 Objective findings

Results of objective experiments show similar frequency response characteristics across the useable range between the impulse responses taken at the venue and those measured within the simulation. (Figure 5). It is anticipated that any significant differences will be corrected by the room equalisation calibration, yet to be implemented. Accuracy in the replication of directional response is assessed by comparing waveform peaks of A-format impulse response channels. Once time-aligned, there is broad agreement in the spacing of amplitude peaks, although the reflections observed in the original venue appear more defined, with greater variation in amplitude. Analysis of reverberation time at 1/3 octave intervals reveals that the simulated acoustic exhibits a shorter T20 value than the prevailing acoustic conditions within the venue and less variance. (Figure 6). Whilst there is broad proportional agreement in time values between the frequency bands from 630Hz upwards, there is less agreement lower down the spectrum. This requires further scrutiny, and it remains to be determined what bearing room equalisation may have upon reverberation time.

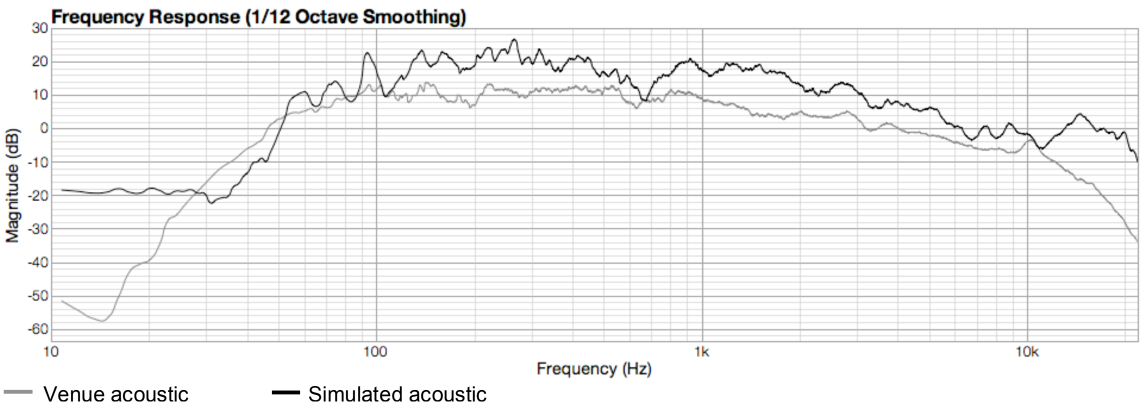

Figure 5: Comparing frequency response of the original impulse response to that measured within the simulation (Thomas Coats Church, position 1, omnidirectional signal).

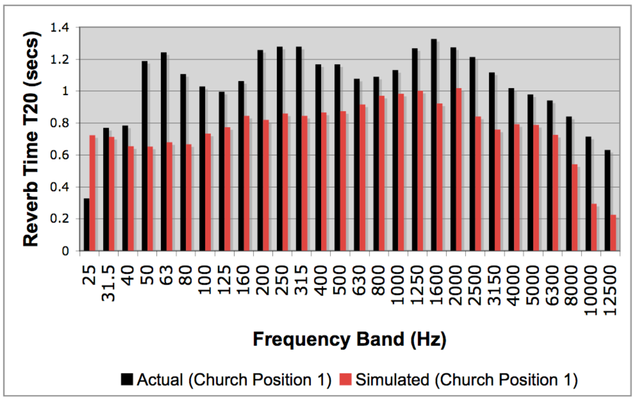

Figure 6: T20 reverb time measured in the venue plotted against that measured within the simulation (Thomas Coats Church, position 1, omnidirectional signal).

5.2 Subjective findings

The mean scores for individual venues and positions are given in Figure 7, below. The recorder performs more strongly than the violin throughout, though the differences are not considerable. Standard deviation scores reveal good agreement between subjects throughout the survey. The Church venue performs consistently well for both instruments and positions, whereas the Thomas Coats Hall exhibits lower mean scores and less agreement between subjects. This is perhaps due to the length and complex timbre of the reverberation tail within Thomas Coats Hall, placing greater demands on the convolution process and making differences more easily perceptible. This may be contrasted with the venue/position with the shortest reverberation time – Brough Hall Stage – which exhibits the strongest performance overall.

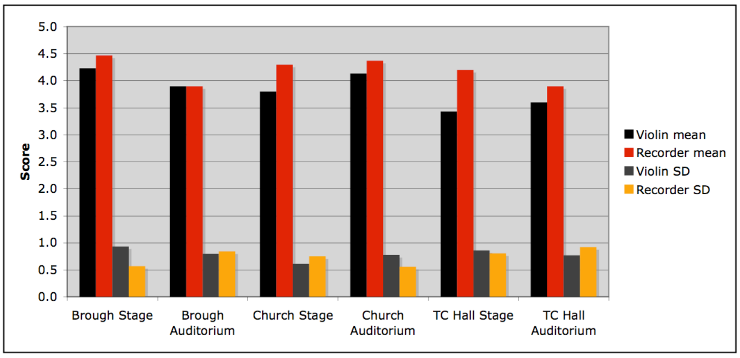

Figure 7: Mean and standard deviation scores by venue position

When analysing the performance of individual parameters it can be seen that Decay Time shows most promise, with highest mean scores and strong agreement between subjects (Figure 8). Comments such as ‘slightly shorter’ and ‘slightly longer’ appear in the survey, whilst other subjects perceived differences in the amplitude envelope of the decay rather than the reverberation time itself. With auditorium aspects requiring greater microphone gain to successfully record instrumental extracts within the venue, background noise was increased, and some subjects suggested that this might account for the relatively ‘clean’ simulated performance *appearing* to have a shorter reverberation tail.

Timbre does not perform quite as strongly, and accordingly there is slightly less agreement amongst those surveyed. Influenced by the convolution algorithm and latency within the plug-in, the mixture of the immediate dry signal and the delayed convolved signal results in destructive interference in the spectrum. A number of subjects perceived a low frequency ‘richness’ in the reverberation tail of the original recorded extracts that seemed lacking in the simulated extracts. This may be related to the aforementioned contribution of microphone gain, but could also point to bandwidth limitations in the convolution processing, since objective measurements revealed a similar finding.

Immersion fairs less well still in regard to mean score and standard deviation. Comments on the differences perceived related mainly to variations in front-back loudspeaker balance and slight changes in the side imaging. Whilst a trend relating to specific venues and receiver positions cannot be identified in this case, it is clear that the mixture and phase interrelationship of the processed and unprocessed signals would again prove to be important.

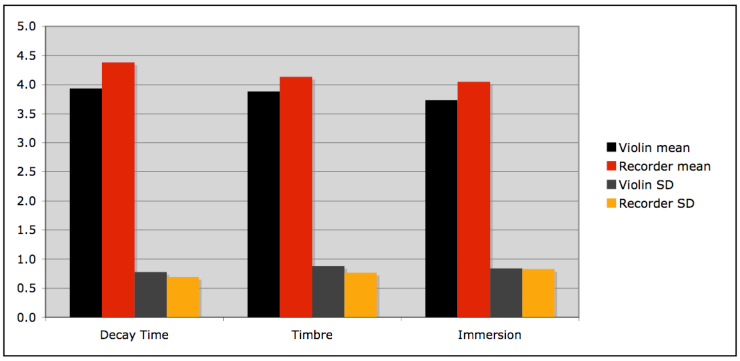

Figure 8: Mean and standard deviation scores by parameter

6 CONCLUSIONS & PROGRESSION

Early objective and subjective tests reveal a respectable reproduction accuracy of the system. Discrepancies in frequency response and reverberation time need to be explored further, and may be rectified via the implementation of additional signal processing for room equalisation and establishing the correct wet/dry balance. System latency needs to be carefully measured too and controlled to ensure that the direct and reverberant sound fields exhibit the correct phase interrelationship. This, in turn will eliminate the discrepancies in the timing of amplitude peaks in directional responses as well as the apparent comb filtering between wet and dry signals at the output of the convolution. Such work will be underpinned by further quantitative assessment.

Plug-ins will be developed to facilitate the additional processing detailed above in addition to ambisonic encoding/decoding and custom multi-channel convolution. Furthermore, the system will be adapted to include modelling of venue occupancy via the post-processing of measured impulse responses.

The fully developed system will be validated via exhaustive qualitative testing involving professional solo musicians preparing for concert performance as well as a broad survey of audio engineers.

7 ACKNOWLEDGEMENTS

The authors wish to thank Robert Geary for ongoing advice, Siobhan Osborne for assistance with the recording of audio extracts, David Gordon for permission to use his composition and Thomas Coats Memorial Church/University of the West of Scotland for access to venues and technical resources.

8 REFERENCES

1. A.C. Gade, 'Investigations of musicians' room acoustic conditions in concert halls. Part I: methods and laboratory experiments, Acustica 69, 193-203'. (1989).
2. K. Ueno and H. Hachibana, 'Cognitive modelling of musicians' perception in concert halls', J. Acoust. Sci. & Tech. 26(2), 156-161. (2005).
3. M. Gerzon, 'Ambisonics in multichannel broadcasting and video', J. Audio Eng. Soc. 33(11), 859-871. (1985).
4. D.G. Malham, 'Spatial hearing mechanisms and sound reproduction', http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm. York (2008).
5. R. Elen, 'Getting Ambisonics Around', http://ambisonic.net/pdf/ambisonics_around.pdf (accessed October 2010).
6. C. Guastavino, V. Larcher, G. Catusseau and P. Boussard, 'Spatial audio quality evaluation: comparing transaural, ambisonics and stereo'. Proc. 13th International Conference of Auditory Display. Montreal (June 2007).
7. K. Ueno, K. Yasuda, H. Tachibana, and T. Ono, 'Soundfield simulation for stage acoustics using 6-channel system', J. Acoust. Sci. & Tech. 22(4), 307-309. (2001).
8. K. Ueno and H Tachibana, 'Experimental study on the evaluation of stage acoustics by musicians using a 6- channel sound simulation system', J. Acoust. Sci. & Tech. 24(3), 130-138. (2003).
9. W. Woszczyk and W. Martens, 'Evaluation of virtual acoustic stage support for musical performance', Proc. Acoustics '08 Conference*,* 1041-1046. Paris (2008).
10. C.C.J.M. Hak and R.H.C. Wenmaekers, 'The effect of room acoustics on the perceived acoustics of reproduced sound', Proc. Inter-noise 37th International Congress and Exposition on Noise Control Engineering, Paper 0274. Shanghai (October 2008).
11. http://www.core-sound.com/TetraMic/1.php (accessed October 2010).
12. British Standards, 'Acoustics – measurement of room acoustic parameters. Part 1: Performance Spaces (ISO 3382-1:2009)', BSI Group. London (2009).
13. Apple Computer, 'Impulse Response Utility User Manual', http://documentation.apple.com/en/impulseresponseutility (accessed October 2010), USA (2009).
14. A. Farina, 'Simultaneous measurement of impulse response and distortion with a swept-sine technique', Proc. 108th Audio Eng. Soc. Convention, Paper 5093. Paris (2000).
15. F. Stan, J. Embrechts and D. Archambeau, 'Comparison of different impulse response measurement techniques', J. Audio Eng. Soc. 50(4), 249-262. (2002).
16. http://www.cockos.com/reaper (accessed October 2010).
17. http://www.radio.uqam.ca/ambisonic/b2x.html (accessed October 2010).
18. http://mcgriffy.com/audio/ambisonic/vvmicvst (accessed October 2010).