Proceedings of the Institute of Acoustics

BRINGING THEATRE SOUND TO THE DESKTOP

Michael Smyth Smyth Research Ltd, Bangor, UK

1 INTRODUCTION

A novel binaural capture and reproduction algorithm is described, termed SVS, that can recreate the acoustics of loudspeaker sources in a real auditorium using normal stereo headphones.

Binaural recording and reproduction techniques aim to capture or recreate the audio signals at each ear. These complementary techniques are commonly used to both predict the acoustics of virtual auditoriums, using modeled room impulse response data, and to assess the acoustics of real auditoriums, using measured data typically captured from a dummy-head microphone.

A major limitation of using dummy-head recorded data is the lack of individualisation, one outcome of which is a person-to-person inconsistency in the spatial and tonal accuracy of the final headphone-rendered audio[1]. The SVS algorithm resolves this problem by capturing the binaural room impulse response of the actual listener to each loudspeaker in an auditorium, thereby accurately recreating, for that individual, all the acoustic characteristics of the speakers in the measured auditorium. In addition, by measuring the impulse responses of the listener at three different head orientations, the SVS algorithm also accommodates head-tracking during the audio rendering stage, the tracking operating within a restricted but still useful range of rotational head movements. The algorithm also includes means to measure and compensate for the non-flat frequency response of each individual to different headphones.

In its current DSP hardware implementation the algorithm is able to recreate simultaneously up to eight virtual loudspeaker sources in the measured auditorium, each virtual loudspeaker being spatially and tonally accurate to the original. When rendering over headphones, head-tracking anchors the virtual speaker sources to their real fixed positions for listener head rotations of up to +/- 30^0 around centre.

2 BINAURAL ROOM VIRTUALISATION

Binaural audio is a recording and reproduction technique that uses two ‘ear’ signals to re-create natural 3-dimensional hearing. Binaural room virtualisation can be defined as a technique that aims to re-create virtual sources (normally loudspeakers) in a reverberant environment, using headphones for delivering the sound to the listener. The technique requires information about how the loudspeakers, the environment and the listener, all affect the audio signals before they are finally heard - information that can be determined from the binaural room impulse response[2,3].

Binaural room impulse response (BRIR) data for each loudspeaker source can be modeled within a computer simulation[4], or measured directly, generally with dummy-head microphones in a real auditorium[5]. To create virtual loudspeakers the measured (or modeled) BRIR for each loudspeaker is convolved with the appropriate audio input signal for that speaker location. The convolved signals for all the virtual speakers are finally summed together and output as a 2-channel 'binaural' signal for playback over headphones.

Dynamic head movements of the listener during reproduction can be accounted for by measuring and storing the BRIR of the dummy head for each loudspeaker at many different head orientations, typically separated by between two and five degrees. A headphone-mounted head-tracker then dynamically interpolates between the nearest sampled BRIR filters, either before or after convolution with the audio signals, resulting in a smooth transition as the listener rotates their head. These core techniques are found, for example, in the well known Binaural Room Scanning system developed by IRT and Studer about ten years ago[5].

Nevertheless, a major issue for accurate binaural reproduction is the requirement for individualised binaural room impulse responses[1]. Over the years it has become accepted that individualisation, whilst necessary for authenticity, is not practical, and that generalised BRIRs, from dummy-heads for example, are therefore more appropriate for commercial applications. However the opposing view offered here is that a binaural reproduction system will only become commercially successful when it can demonstrate sufficient accuracy, and that this can only be achieved using personalised BRIR data. From this perspective the real problem to be solved is how to practically include personalisation in binaural room virtualisation, since this typically requires a high degree of user-customisation, one that would tend to limit its commercial appeal.

Further problems with using headphones to render the binaural audio are the lack of dynamic spatial cues, the need for personalised headphone equalisation filters, and lack of sub-sonic frequencies.

3 DESIGN AND FEATURES OF THE SVS ALGORITHM

The key novel feature of the SVS algorithm is its ability to measure and use personalised BRIR data within a real-time head-tracked convolution system for headphone rendering. The SVS algorithm therefore comprises three main parts: a binaural room impulse response and headphone EQ measurement system, a streamlined multichannel convolution engine, and a simple, low-cost head-tracking system[6].

3.1 PRIR: Personalised Room Impulse Response

An important feature of the SVS algorithm is the capture of personalised binaural room impulse responses (termed PRIR). These are measured from each loudspeaker in a real listening environment, using miniature microphones placed at the entrance of the listener's ear canals, in the blocked-meatus configuration. The loudspeaker excitation signal, generated from the SVS algorithm, is a swept sine wave of 3s or 12s length, with optional multi-sweep averaging[7].

The impulse response measurement procedure is relatively simple and was specifically designed to allow users to measure themselves in the sweet-spot of a 5.1-channel surround-sound studio. A sparse set of binaural impulse responses for all active loudspeakers are recorded at user-head orientations of approximately -30^0, 0^0 and $+30^0$ azimuthal angle, corresponding to the user orientating their head to face the left, centre and right loudspeakers. A simple extension of this technique uses the head-tracker to determine the three head orientations, for situations where the speakers are not visible or their angular positions are not known.

The three 'look-angles' chosen allow simple rotational head-tracking to be accomplished by interpolation between the binaural data sets from each head position. Essentially head-tracking operates for angular head movements within the scope of the left and right stereo loudspeakers. This may appear somewhat restrictive, but is usually adequate for professional applications where the normal monitoring position is looking straight ahead with only minor amounts of head movement.

The final PRIR data set consists of binaural room impulse responses for up to eight loudspeakers at three head orientations, and is usually measured in about six minutes using the 12s sweep, or under two minutes using the 3s sweep.

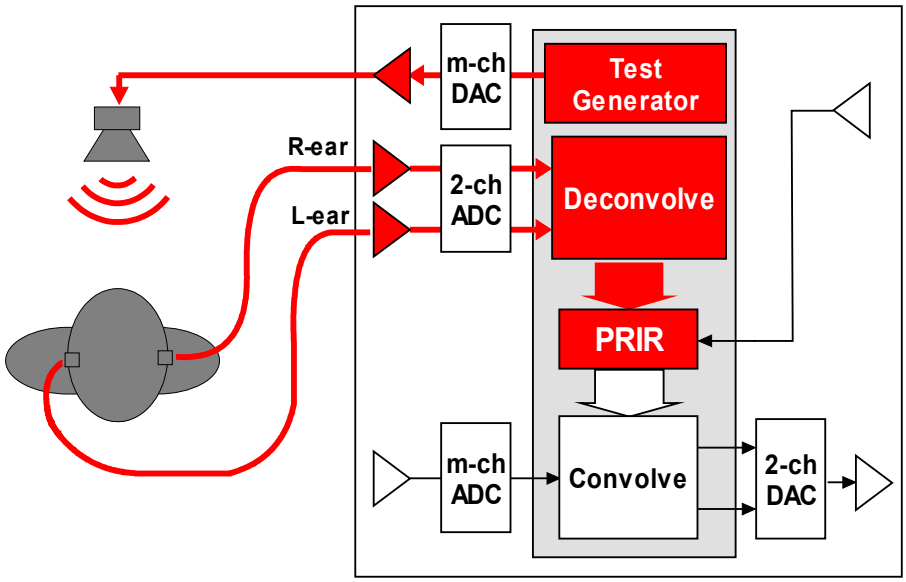

Figure 1: Integrated test signals and procedures within the SVS algorithm for acquiring personal binaural room impulse response (PRIR) data.

3.2 Headphone Equalisation

Headphone equalisation is needed to compensate for the filtering effect of the listener's outer ear, or pinna, when the virtualised audio is finally presented to the listener over headphones. SVS uses a two stage procedure. The first stage produces a filter automatically, and the second stage allows for manual adjustment of these filter coefficients.

In the first stage the same miniature microphones are placed in the same blocked-meatus location of the listener's ear canal and the headphone-to-pinna response for each ear is recorded using a short sine sweep excitation signal emitted from each headphone driver. The headphone equalisation filters for an individual user are then automatically generated and stored. These filters essentially invert the measured headphone-to-pinna response and are specific to each user and headphone model. The degree of inversion is restricted and can be adjusted by the user in three frequency bands. This automatic headphone equalisation measurement is normally conducted immediately after the PRIR measurement and takes less than one minute to complete.

If required, the auto-generated headphone equalisation filters can be manually modified to take account of the filtering effect of the listener's ear canal. This is accomplished by directly comparing the volume of real and virtual loudspeakers when listening to narrow-band pink noise signals. Manual adjustment is an iterative task, and normally takes around ten to fifteen minutes to complete.

All the excitation signals and measurement procedures for generating the PRIR and headphone EQ data are integrated within the SVS algorithm, and the operations are almost fully automated through the use of audible cues to guide the user at each stage.

3.3 DSP Convolution engine

Each measured PRIR contains all the information necessary to virtualise a single loudspeaker, placed anywhere in the auditorium, for one particular head orientation. The three measured sets of PRIRs will therefore virtualise all the active speakers in three head positions, looking left, centre and right. This allows dynamic head-tracking over the left-to-right speaker range by interpolating a new set of PRIRs between the three measured data sets.

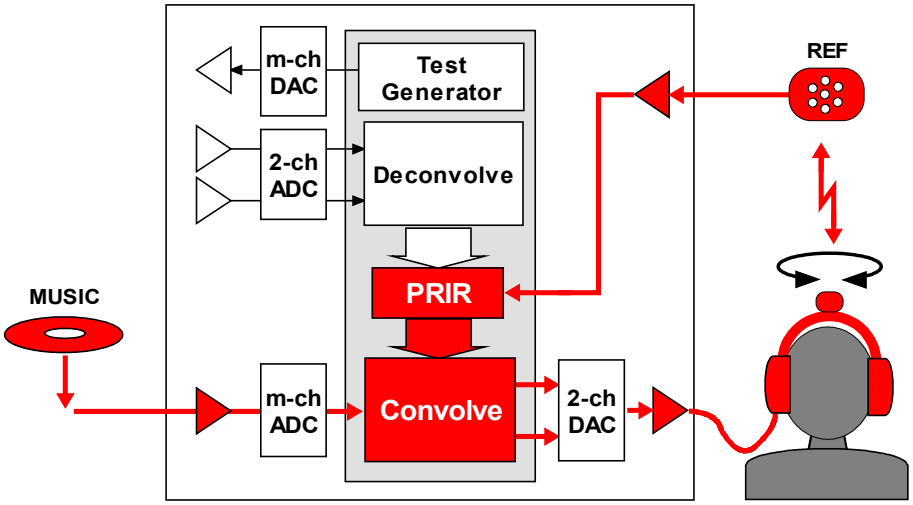

Figure 2: Convolution of PRIR data with up to eight channels of incoming speaker source audio, generating a binaural headphone signal. PRIR interpolation and ITD insertion is controlled by a headphone-mounted optical head-tracker.

Interpolation between the PRIR data is accomplished by removing the inter-aural time difference (ITD) between each left-ear, right-ear impulse pair, interpolating between the measured PRIRs, driven by the measured head azimuthal angle, and then re-inserting a new appropriate ITD at a sub-sample resolution, again driven by the measured head angle. This is a relatively simple but robust interpolation technique.

After ITD insertion each binaural filter pair is convolved with an individual speaker signal, and the convolved signals, to a maximum of eight, are summed to a single pair of binaural signals. Finally, before being output, the signals are equalised, using the left-ear and right-ear filters calculated from the measured headphone-pinna impulse responses.

The final output is a personalised binaural signal suitable for presentation over headphones, containing up to eight virtual speakers. The position of all the virtual speakers are anchored to their true, external position, by tracking the orientation of the listener's head and dynamically changing the virtualised sound field. Using a tilt detector in the head-tracker it is possible to directly compare the real sound field (headphone off the head) with the virtual sound field (headphone on the head), at any head orientation between the left and right look-angles.

3.4 Personalised head-tracking

Headphone based virtualisation, even using individualised binaural data, can still cause confusion, particularly front-back reversals, if the virtual speakers rotate in lock-step with the subject's head.

Head-tracking can remove this confusion by continuously tracking and measuring the rotational angle of the user's head. Within the SVS algorithm the head-angle is used to calculate an interpolated PRIR data set appropriate for this new head orientation, and this also adjusts the inter-aural time difference (ITD) within each binaural signal pair for each of the active speakers. Finally all the binaural signals for convolved with the source signals and the outputs are summed for headphone presentation. The positions of all the virtual speakers are now dynamically locked to their real external position, for all head orientations that fall within +/-30^0 around centre.

This head-tracking ability is integrated within the core SVS algorithm, but it can be easily disabled in which case SVS reverts back to operating at a head-angle of 0^0, i.e. looking centre. The head-tracking algorithm operates asynchronously to the incoming head-angle data, up to a maximum rate of 182 Hz, and has a total latency of approximately 16ms, measured from the initial head movement to the compensated headphone output.

3.5 Current hardware implementation: Realiser A8

The Realiser A8 (Figure 3) is the first implementation of the SVS algorithm, and was designed primarily for consumer applications with unbalanced analogue inputs and outputs. The portable processor can store internally up to 64 PRIR data-sets and 64 HPEQ filters, and can store others on an SD card.

The floating point DSP within the A8 processor is capable of virtualising and tracking eight full-bandwidth speakers, in any position, with a maximum convolved reverberation time of 850ms per speaker. The internal processing latency is approximately 16ms, from the time of analogue audio input to virtualised headphone output.

Figure 3: Smyth SVS Realiser A8, an 8-channel real-time headphone virtualisation processor. Shown with optical head-tracker mounted on stereo headphones, and the reference set-top IR transmitter.

3.6 Optical head-tracking device

Figure 4 illustrates an example of a simple headphone-mounted head-tracking device that was specifically designed for operation with the SVS algorithm. Using an external infra-red (IR) light source as the 0^0 reference plane, it can detect angular rotations of approximately 0.25^0 over a range of $+/-60^0$, with a latency of approximately 5ms, and operates at distances up to 6m from the IR reference.

3.7 Key features of SVS Realiser A8

1. Any environment to which the user has access can potentially be captured.
2. Loudspeakers can be in any location with respect to the subject listener.
3. Multiple measurements can be made at different seating positions for comparison.
4. Can be used with any type of stereo headphone.
5. Can also upload modeled BRIR data from a room acoustics prediction program.

Current limitations - with work-around

1. Maximum of eight loudspeakers per Realiser A8 – can use multiple 8-speaker processors simultaneously and externally mix the analogue headphone outputs to a single headphone signal. All the processors can be driven simultaneously by one head-tracker signal.
2. Lack of tactile response for subsonic frequencies – can use an external vibration plate driven by the low-pass filtered audio signals. The current Realiser A8 has a dedicated low-pass filtered tactile output that can be generated from the input or output signals .

Current limitations - no work-around

1. Assumes a linear, time-invariant system - non-linear characteristics are not recreated.
2. Must use stationary loudspeakers as sources.
3. Maximum measured reverberation length of 850ms.
4. For accurate results the listener must be able to be measured on-site.

4 APPLICATION TO THEATRE SOUND REPRODUCTION

Since SVS gives users the ability to easily measure their own binaural impulse response to real loudspeakers in a real environment, it has the potential to replicate loudspeakers in almost any auditorium even for novice users. To date successful measurements have been made in film dubbing stages and screening rooms in LA, New York and San Francisco, a large movie theatre (Egyptian Mann) in Hollywood, and the state opera house in Hamburg.

4.1 Example: Hamburg State Opera

Jochen Schulz, an acoustics engineer working in the Sound Department of the Hamburg State Opera, has the task of balancing the sound of each production, so that accompanying singers, instruments and choir are well matched to the live acoustic orchestra. If space on or behind the stage is restricted, the choir or small instrumental group must be placed in a separate room or building and the signals transmitted live to loudspeakers on the stage. Occasionally some music must be prerecorded and then played back from the loudspeakers during the show. Also, for non-operatic music shows such as pop music, more forceful sound reinforcement is generally required.

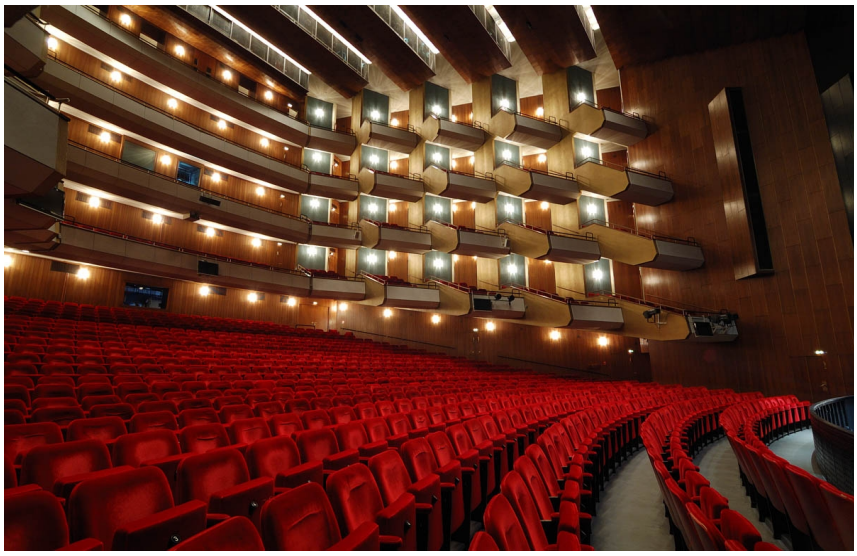

Figure 4: Hamburg State Opera Hall

In addition to using the Realiser as a substitute for a professional audio control room (a typical application) Jochen was also interested in measuring the sound stage of the actual opera hall and using this to adjust the audio playback of the reinforcement speakers for new opera pieces. Normally this must be done while the stage and lighting crews are working in the hall and this invariably restricts the amount of useful audio monitoring time. Also, as more directors are now starting to take advantage of the installed surround, ceiling and balcony speakers for special effects, careful monitoring and rehearsal of the sound production is required. To solve these problems Jochen envisaged capturing eight of the main loudspeakers (Figure 5) from different seats in the public area, then using these measurements to do most of the audio preparation work using just the audio mixing desk and Realiser.

Figure 5: Main front reinforcement loudspeakers, Hamburg State Opera. Others are situated in the roof, side walls and balcony.

Initial results have been very encouraging and, according to Jochen, are subjectively 'really precise'. The 850ms limitation of the measured reverberation time (Figure 6) is not normally apparent, apart from with percussive sounds, and its influence can be reduced further by tapering the reverberation tail of the measured PRIR data to zero rather than having an abrupt cut-off. The eight-speaker limitation has so far not been an issue since most productions only use a sub-set of the more than twenty available speakers. This project is still in the early stages of development but Jochen is very optimistic about its potential to help him improve the quality of sound reproduction.

5 CONCLUSION

Whilst the SVS algorithm was designed for measuring small listening and control rooms, it has proven capable of capturing and accurately simulating much larger auditoriums. SVS allows users to independently measure their own binaural room impulse responses in almost any environment, and to then use this data to recreate the acoustics at a desk-top audio workstation. Preliminary results from measurements taken with the Smyth SVS Realiser A8 in the main hall of the Hamburg State Opera, indicate that the simulation is very accurate and should readily allow off-line audio production and preparation work to be undertaken.

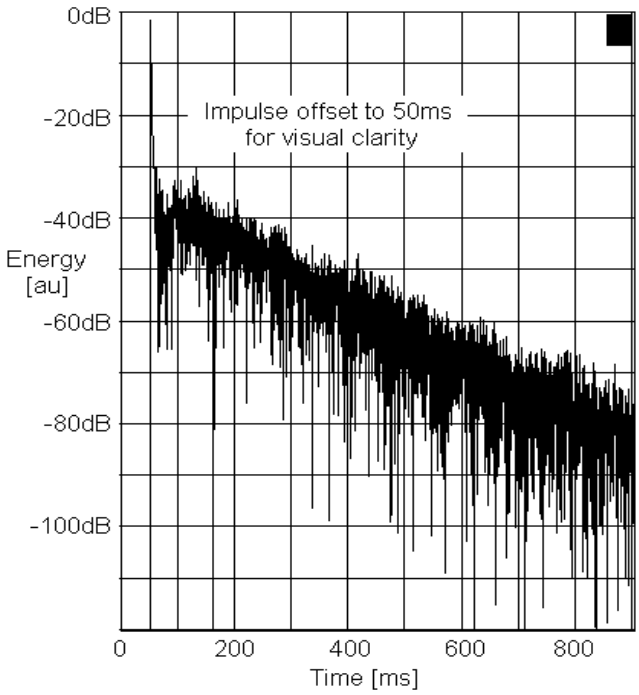

Figure 6: Energy decay curve, Hamburg State Opera Hall (main front lower speaker, left ear signal, looking centre, row 11, centre seat)

6 REFERENCES

1. W.G. Gardner. “Spatial Audio Reproduction: Toward Individualized Binaural Sound.” The Bridge Vol. 34 No. 4: 37-42

2. W.M. Hartmann and A. Wittenberg. “On the externalization of sound images” Journal of the Acoustical Society of America 99 (6): 3678–3688

3. F.L. Wightman and D.J. Kistler. “Headphone simulation of free-field listening I: stimulus synthesis” Journal of the Acoustical Society of America, 85(2) 1989a.

4. A. Persterer. “Binaural Simulation of an Ideal Control Room for Headphones Reproduction” 90th Convention, Audio Eng. Soc., Paris 1991, Preprint 3062

5. U. Horbach, A. Karamustafaoglu, R. Pellegrini, P. Mackensen, G. Theile. “Design and Applications of a Data-based Auralisation System for Surround Sound” 106th Convention, Audio Eng. Soc., Munich 1999 Preprint 4976

6. S.M.F. Smyth et al. Smyth SVS. “Headphone Surround Monitoring for Studios” Presented at the 23rd UK Conference Audio Eng. Soc., Cambridge 2008

7. A. Farina. “Simultaneous Measurement of Impulse Response and Distortion with a Swept-sine technique” 108th Convention, Audio Eng. Soc., Paris 2000 Preprint 5093