Auralisation of Stage Acoustics for Large Ensembles

Daniel Jimenez, Nicholas Vandenberg, Luis Miranda, Manuj Yadav, and Densil Cabrera

Faculty of Architecture, Design and Planning, University of Sydney, NSW, 2006, Australia

ABSTRACT

Computational room acoustics models are frequently used to derive room acoustical parameters, and also for auralisation of sound from stage to audience area of auditoria. Auralisation of sound on stage is much less common. This paper considers how auralisation can be implemented for a musician within a large ensemble. Using a computer model of an auditorium, impulse responses are derived from various source positions on the stage to the auralised position for the musician (also on the stage). For the playback, anechoic recordings of individual instruments within a small or large ensemble are convolved with the appropriatiately synchronised and spatialised impulse responses. The room-reflected sound for the musician's own instrument or voice is auralised separately using a real-time convolution system previously developed for auralising the sound of one's own voice in rooms. This paper describes the integration of these processes, which are illustrated using a case study.

INTRODUCTION

Computational room acoustic models used for auralisation of a stationary position on stage to a stationary position in the audience are used regularly when analysing the acoustic characteristics of that space. However, the use of computer models to auralise the acoustic conditions that a performer experiences on stage is less common. This paper describes an auralisation system for an ensemble scenario, by combining the real-time auralisation of a musician's instrument or his/her own voice on stage and the auralisation of accompanying musicians. The system is further enhanced with a head-tracking system. In simple terms, the system could be thought of as a karaoke system, where the performer plays his/her instrument or sings in the same auralised room as the accompanying musicians.

Analysing the influence of variable room acoustics on the ability of musicians to play in ensemble has proven useful to derive parameters that relate to the musicians' impressions on stage (Gade 1989), and has led to more controlled experimentation in recent years. For instance, recent systems have used computer generated stimuli (Guthrie 2008) or pre-recorded impulse responses (Ueno & Tachibana 2003) (Ueno et al. 2005) to manipulate the acoustic environment presented to the musicians. However, these studies have mainly focused on the ability of a very limited number of musicians to play in ensemble. Some of the limiting factors in ensemble auralisation are the computational requirements for real-time interactioin and the need for special isolated spaces (usually independent anechoic chambers) to accomodate the performers. In the auralisation system proposed in this paper the focus is on large ensembles, and while the ability to analyse two-way communication between pairs of musicians is not supported, the ability to understand musicians in a larger context is incorporated.

The perception of the sound of one's own voice, i.e. autophony (Lane et al. 1961), has received some attention in recent years (Brunskog et al. 2009; Pelegrín-García, Smits, et al. 2011b; Pelegrín-García, Fuentes-Mendizábal, et al. 2011a). From these studies, new parameters have been derived to describe the effect of the room acoustics on the autophonic level. However, it is hard to generalise the results of these studies to singing conditions or listening to one's own instrument as these studies have been solely focused on the study of speech tasks.

There are studies in the literature where the preferences of singers for various acoustic conditions have been considered. However, the majority of these studies have been limited to the study of the influence of one reflection (Marshall & Meyer 1985) and limited to small choral ensembles (Noson et al. 2000; Noson et al. 2002).

The system described in this paper combines the room simulation system for perceiving autophony described by the authors (Yadav, Cabrera & Martens 2012b) with the simulation of a large ensemble, derived from the calculations of a computational room acoustic model. This allows the study of a large number of different room acoustical conditions provided by the flexibility of computational room acoustic calculations as well as the availability of anechoic recordings (Patynen et al. 2008) to simulate singers/instruments in large ensembles.

The autophony room simulation system has been successfully implemented for tasks involving the perception of one's own voice in rooms (Yadav et al. 2011). Following this, the system proposed in this paper is solely concerned with the auralisation of rooms for vocal performance as part of an ensemble performance. There are potential modifications of this system for instruments where the performers' ears are close to the instrument's main radiation lobe (e.g. violin); however, these are not considered in this paper. The system proposed in this paper includes the auralisation of an orchestra in a concert hall where the focus of the room acoustic calculations is a point representing one of the musicians within an ensemble. All the room acoustic calculations were performed in ODEON and processed for playback using MATLAB. The process is detailed with a sample case in the following sections. Finally, an alternative to the auralisation system of the orchestra is introduced to explore the possibilities of using a high density hemispherical loudspeaker array similar to the systems described by Zmölnig, Ritsch, & Sontacchi, (2003) or Zotter, Pomberg, & Noisternig, (2010). This procedure would allow a realistic spatial rendering of the orchestra in the concert hall, which is integrated with the component for simulating autophonic perception (Yadav, Cabrera and Martens, 2012b).

Each of these components is elaborated below.

THE AUTOPHONY SIMULATION SYSTEM

The autophony simulation system has been described previously (Yadav, Cabrera & Martens 2012b) and the system based on computer room model calculations has been described in detail by Yadav et al. (2012a). For this paper a brief overview is included in order to explain the interaction of this system with the ensemble auralisation.

The autophony system operates using a set of near-ear loudspeakers that reproduce the room reflected sound of one's own voice convolved in real time with an Oral Binaural Impulse Response (OBRIR) (Cabrera et al. 2009). The OBRIR is selected according to the users' head rotation information, obtained from a head tracking system. As the user rotates his/her head, the OBRIRs are cross-faded providing a seamless transition between head positions. The use of ear loudspeakers allows the user to receive the direct sound from his/her mouth with little effect on this direct sound path. A reflector is also positioned on the floor of the anechoic room (in which the simulation is realised) to produce a floor reflection. Since the direct sound and floor reflection are not included in the system's computer auralisation, the delay incurred by the convolution process does not have an impact on the timing of early reflections because the first 7 ms of the OBRIRs can simply be truncated (allowing for 7 ms of processing latency).

The autophony simulation system is designed to operate in Max/MSP, hosting a SIR2 convolution plugin on a Windows based computer. The AD/DA converter used is a RME ® ADI-8 QS unit with 48000 Hz sampling rate and 32-bit quantization in a 1-in/2-out configuration. The headset microphone used for the vocal input is a DPA 4066 positioned at a distance of 7 cm from the centre of the lips on the right side of the face to minimise the proximity effect that occurs when air blown from the mouth passes the microphone. The ear loudspeakers used are a pair of AKG K1000 that are positioned near the ears..The head tracker used is a Polhemus Fastrak® unit with a refresh rate of 5 Hz, which is used is update the talking/listener's head position.

The method for deriving impulse responses from computer room models using higher order Ambisonics (HOA) principles presented in (Yadav, Miranda Jofre, Cabrera & Martens 2012a), is used for deriving the OBRIRs in this paper. Briefly, in this method, room impulse responses are produced using computational room acoustic calculations. The source is modelled with dispersion patterns that correspond to each of the HOA components up to fifth order. The source can then be beamformed by combining the HOA source components; in this case it is beamformed to a directivity pattern that resembles the directivity of a Head And Torso Simulator (HATS). The received impulse responses contain complete directional information, which is then encoded into HOA signals. The HOA signals from the receiver side are then beamformed to a predefined number of directions evenly spaced around a sphere. The beamformed signals are then convolved with head-related impulse responses (HRIRs) for these discrete directions. The beamforming on source and receiver sides and HRIR convolution processes are repeated for the required number of head position angles.

ENSEMBLE AURALISATION SYSTEM

Computer room model creation and verification

The room acoustic calculations were performed using the ODEON software (Rindel 2000). This software has been successfully tested previously to provide reliable results for room acoustic calculations (Vorländer 1995; Bork 2000; Bork 2005). This software has also been extensively used in auralisations tasks (Choi & Fricke 2006; Favrot & Buchholz 2010; Vigeant et al. 2008; Vigeant et al. 2011).

The challenge in the present case is to create a realistic and acoustically accurate auralisation of an orchestra. In order to create a realistic auralisation, it was decided to base our computer model on a real performance space (Olvashallen) that has a good reputation of acoustics for opera performances and also noteworthy architectural features to aid its acoustic performance (specifically, a removable orchestra shell and existing floating panels).

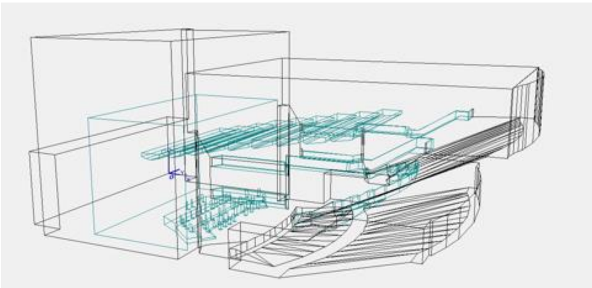

Figure 1: Olvashallen computer model in ODEON.

The model was built based on the published plans and surface material characteristics (Beranek 2004), and is illustrated in Figure 1. The acoustical parameter used for matching the model to the real space was reverberation time. The compared results for the published average unoccupied condition values (Beranek 2004) and the results obtained from the model are shown in Figure 2.

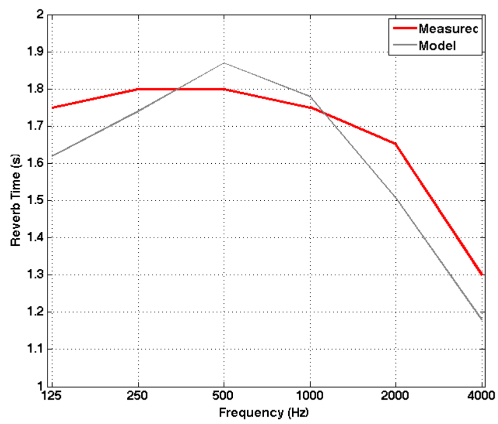

Figure 2: Comparison of reverberation times from modelled and published results.

The orchestra was modelled with individual sources representing each of the instruments in an orchestra. The location of the instruments was based on a typical American seating arrangement as described in (Meyer 2009) and shown in

Figure 3. The directivity of the instruments was also included in the model calculations. The directivities were obtained from data presented in (Patynen & Lokki 2010) and exported into ODEON.

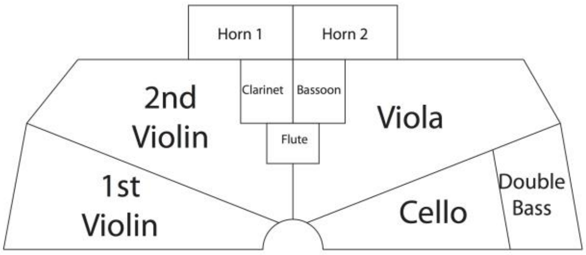

Figure 3: Orchestra layout used in simulation.

An important inclusion for the room model is the additional acoustical absorption/obstruction from musicians on stage, which affects the direct path between musicians. To model this, the musicians were modelled as two orthogonal surfaces located underneath each of the sources representing individual musicians. The findings of (Dammerud 2010) provided the data for absorption, transmission and scattering included as properties for these surfaces.

Room calculations were performed using the model described above. Responses were calculated between each individual instrument and the location of the singer on stage. The settings in ODEON were set based on the recommended settings for precision calculations. The impulse response resolution was changed to 1 ms to increase the available data points for the calculations described below.

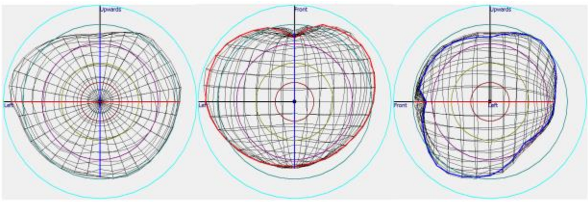

Figure 4: Example of directivity pattern in ODEON. Violin pattern shown.

Deriving spatial playback signals based on HOA processing

One of the main features of the proposed auralisation system is the inclusion of a head-tracking system. The head tracking system follows the convention of the autophony head-tracking system, i.e. the system has a range of -60° to 60° in the horizontal plane with a resolution of 2° (Yadav, Cabrera and Martens, 2012b). If we were to derive impulse responses for the auralisation of each individual source for all the necessary angles, the required number of room acoustic calculations would be large and the process time-consuming. In order to reduce the required calculations a similar approach as proposed in (Favrot & Buchholz 2010) for room renderings for playback and in (Yadav, Miranda Jofre, Cabrera & Martens 2012a) for room impulse response derived from HOA processing is considered.

The data from the room acoustic calculations provided by ODEON can be accessed in text-based format and provides complete directional and temporal information. This is true for the early reflection information contained in the reflectogram text export and the late energy histogram contained in the decay curve text export.

We then use the directional information of arrival to decompose the soundfield at the location of interest into its HOA components, resulting in a set of impulse responses for each HOA component. Under HOA, the soundfield is decomposed into its spherical harmonic components which can be later used to resynthesise the soundfield at the listener position (Daniel et al. 2003). The process outlined in this part of this paper was achieved by processing the text output of ODEON in MATLAB.

In our case we are calculating the decomposition of the soundfield into its HOA components at the location of the auralisation. The decomposition of the soundfield into its HOA components is achieved by weighting the reflection intensity at a time instance by its corresponding spherical harmonic function. The spherical harmonic function weight is given to an azimuth (θ) and elevation (φ) angle-pair and described by the following equation:

$$Y_n^m = (-1)^m \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta)e^{im\varphi} \quad (1)$$

where m is the spherical harmonic order, n is the spherical harmonic degree and $P_n^m(\bullet)$ is the associated Legendre function (Arfken 1985).

The process of creating HOA impulse responses is done for every source to be included in the auralisation. The HOA impulse responses for each source are then convolved with an anechoic recording of the relevant instrument, providing a set of HOA signals. In our case we are using the anechoic signals recorded by (Patynen et al. 2008). For this project the recording of Donna Elvira from the opera Don Giovanni by W.A. Mozart is being used. This piece was selected as it was the only piece to include a singer within the available multitrack anechoic music recordings.

In order to add realism to the auralisation, each instrument has been modelled as a separate source. However, the available recordings only include one recording for each part for each instrument. To overcome this, each instrument section that requires more than one single instrument has been processed in a manner akin to chorus processing used in digital audio effects commonly used in recording environments (Zolzer 2011). The chorus effect is obtained by applying delay modulation on the recorded signal; the modulation signal is a low frequency random signal.

Using the HOA signals corresponding to each instrument, we use beamforming to a set of sources in space that are paired to an HRIR associated with a specific direction. This process is similar to the decoding process usually encountered in the decoding of HOA signals for playback from discrete loudspeakers. The signals are then convolved with a set of measured HRIRs to create a binaural reproduction of the soundfield as described in (Song et al. 2008). Using HOA decoding allows us to create a new set of binaural signals simulating head rotations by changing the set of HRIRs being used for the convolution process. A spatial interpolation technique as described in (Supper 2010) is used to obtain new HRIRs for the missing measurement location. The interpolation technique is based on weighted interpolation in the time domain.

The last step in creating the spatial ensemble auralisation for a single angle is to sum all binaural renderings of the individual instrument. Up to this stage there has been no gain correction in any of the processes involved in producing the signals. Since this is the case and all processes are linear, the obtained signals will have correct level relationships between them.

Lastly, to create a complete set of ensemble signals for the head tracking system, a new beamforming and convolution with HRIR process is followed for each angle required. The complete process is shown in Figure 5.

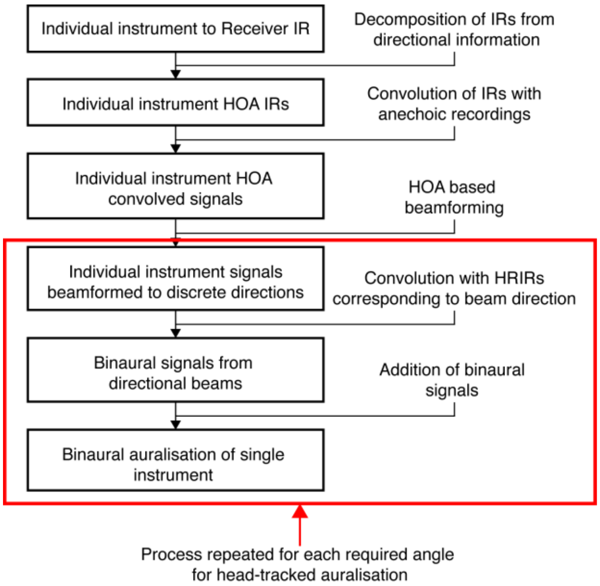

Figure 5: Overview of ensemble auralisation with head-tracking process.

INTEGRATING THE AUTOPHONY AND ENSEMBLE AURALISATION SYSTEMS

The final procedure to create the ensemble auralisation system involves combining the autophony system and the ensemble auralisation system. As stated above, the system responds to head movement obtained from the head-tracking system worn by the user. Both systems rely on individual signals for each of the discrete steps, which correspond to the system's angle resolution. In the case of the autophony system these signals are an OBRIR used to convolve the user's voice; in the case of the ensemble system, a pair of binaural signals corresponding to the room rendering of the ensemble.

The system used has the capacity to read multiple tracks at once, while at the same time performing real-time convolution. The ensemble auralisation is comprised of a multiple track audio file, organised in pairs associated to a specific head angle position. The autophony system uses the same amount of OBRIRs paired to the equivalent head position. For example, if we use the system's convention as it has been previously used (auralisation from -60° to 60° in the horizontal plane in 2° steps), we can expect to use 61 binaural pairs of ensemble signals. The autophony system for the same resolution will use 61 OBRIRs.

The equivalent pairs of stimuli are synchronised so that an OBRIR tied to a specific angle is used to convolve the user's voice at the same time as playback for the equivalent angle reaches the user's ears. A 10 ms fade is applied to both the ensemble auralisation and convolution playback as the user turns his/her head and the OBRIR and ensemble auralisation shifts to the new angle.

ALTERNATIVE TO THE ENSEMBLE AURALIZATION SYSTEM

As an alternative, the set of HOA signals corresponding to each source modeled within the ensemble can be decoded to be played back from a discrete number of loudspeakers instead of the AKGK1000 to arguable achieve a higher spatial fidelity

The proposed array consists of 196 Mod 1 Orb Audio loudspeakers placed in each vertex and strut of a frequency three geodesic dome structure which comprises 5/9th of a spherical array (Kruschke, 1972) of 2.2 meters radius as shown in figure 6. An arrangement of these characteristics could be used as a playback system for the decoded HOA signals as described by Zotter et al. (2010).

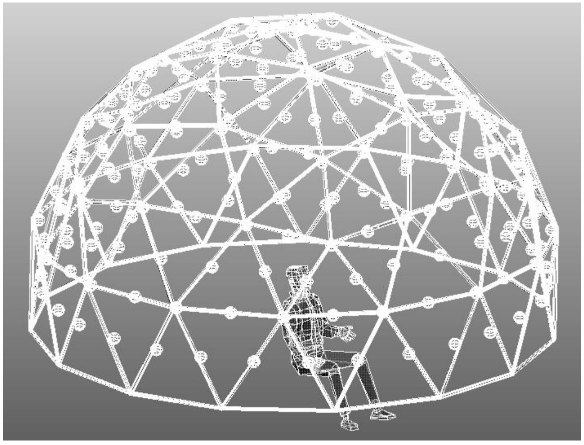

Figure 6: Gesodesic dome structure proposed to configure the high density loudspeaker array.

The use of such system would greatly improve the spatial cue information received by the listener in terms of where each of the sections of the ensemble is located in the simulated soundfield.

The segregation of the systems used to render the soundfield and the autophony simulation allows individual control, whilst maintaining the acoustic conditions so that they can simultaneously produce a more realistic approximation to the environment that would be experienced on stage during a musical performance.

CONCLUSION

An auralisation system that combines an autophony auralisation system and ensemble auralisation system with head tracking based on computational room acoustics calculations has been presented. The system has potential uses in understanding the acoustic characteristics preferred by performers when playing along a large ensemble in an acoustic environment that closely resembles normal performance conditions. The system presented here offers an added flexibility compared to previous systems as it relies on modelling to produce the auralisation stimuli. Future studies will focus on the ability of the system to provide musicians with acoustic environments with varying acoustics based on common acoustic descriptors (reverberation time, clarity index, stage support), as well as acoustic descriptors currently being developed.

ACKNOWLEDGEMENTS

The authors wish to thank Ken Stewart for technical assistance. Craig Jin and Nicolas Epain helped with the concepts relating to HOA. This research was supported under the Australian Research Council's Discovery Projects funding scheme (project DP120100484). This research was also supported by an Australian Postgraduate Award for Manuj Yadav and Luis Miranda. LM is further supported by a CONACYT (Mexico) scholarship. MY is further supported by Faculty of Architecture, Design and Planning Top-up and Jean and Andrew Wong scholarships.

REFERENCES

Arfken, G., 1985. *Mathematical Methods for Physicists*, San Diego, CA: Academic Press.

Beranek, L., 2004. *Concert Halls and Opera Houses - Music, Acoustics, and Architecture* 2nd ed, New York, NY: Springer-Verlag.

Bork, I., 2000. A comparison of room simulation software - The 2nd round robin on room acoustical computer simulation. *Acta Acust United Acust*, 86(6), pp.943–956.

Bork, I., 2005. Report on the 3rd round robin on room acoustical computer simulation Part II: Calculations. *Acta Acustica united with Acustica*, 91(4), pp.753–763.

Brunskog, J. et al., 2009. Increase in voice level and speaker comfort in lecture rooms. *The Journal of the Acoustical Society of America*, 125(4), pp.2072-2082.

Cabrera, D. et al., 2009. Binaural measurement and simulation of the room acoustical response from a person's mouth to their ears. *Acoustics Australia*, 37(3), pp.98–103.

Choi, Y.-J. & Fricke, F., 2006. A comparison of subjective assessments of recorded music and computer simulated auralizations in two auditoria. *Acta Acustica united with Acustica*, 92(4), pp.604–611.

Dammerud, J.J., 2010. *Stage Acoustics for Symphony Orchestras in Concert Halls*, University of Bath, England. PhD Thesis.

Daniel, J., Nicol, R. & Moreau, S., 2003. Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging. In AES 114th Convention. Amsterdam, The Netherlands.

Favrot, S. & Buchholz, J.M., 2010. LoRA: A loudspeaker-based room auralization system. *Acta Acustica united with Acustica*, 96(2), pp.364–375.

Gade, A.C., 1989. Investigations of musicians' room acoustic conditions in concert halls. Part I: Methods and laboratory experiments. *Acustica*, 69, pp.193–203.

Guthrie, A.E., 2008. *Using Telepresence to Evaluate the Relevance of Acoustic Parameters for Musician Communication in Indeterminate Music*, Rensselaer Polytechnic Institute, Troy, NY. M.Phil Thesis.

Kruschke, D. (1972). *Dome Cookbook of Geodesic Geometry.*

Lane, H.L., Catania, A.C. & Stevens, S.S., 1961. Voice level: autophonic scale, perceived loudness, and effects of sidetone. *Journal of the Acoustical Society of America*, 33(2), pp.160–167.

Marshall, A.H. & Meyer, J., 1985. The directivity and auditory impressions of singers. *Acta Acustica united with Acustica*, 58(3), pp.130–140.

Meyer, J., 2009. *Acoustics and the Performance of Music* 5th ed, New York, NY: Springer-Verlag.

Noson, D. et al., 2002. Melisma singing and preferred stage acoustics for singers. *Journal of Sound and Vibration*, 258(3), pp.473–485.

Noson, D. et al., 2000. Singer responses to sound fields with a simulated reflection. *Journal of Sound and Vibration*, 232(1), pp.39–51.

Patynen, J. & Lokki, T., 2010. Directivities of symphony orchestra instruments. *Acta Acustica united with Acustica*, 96(1), pp.138–167.

Patynen, J., Pulkki, V. & Lokki, T., 2008. Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica*, 94(6), pp.856–865.

Pelegrín-García, D., Fuentes-Mendizábal, O., et al., 2011a. Equal autophonic level curves under different room acoustics conditions. *The Journal of the Acoustical Society of America*, 130(1), pp.228-238.

Pelegrín-García, D., Smits, B., et al., 2011b. Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129(4), pp.1981-1990.

Rindel, J.H., 2000. The use of computer modeling in room acoustics. *Journal of Vibroengineering*, (3), pp.41–72.

Song, W., Ellermeier, W. & Hald, J., 2008. Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise. *Journal of the Acoustical Society of America*, 123(2), pp.910–924.

Supper, B., 2010. Processing and improving a head-related impulse response database for auralization. In 129th Convention of the Audio Engineering Society. San Francisco, CA.

Ueno, K. & Tachibana, H., 2003. Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system. *Acoustical science and technology*, 24(3), pp.130–138.

Ueno, K., Kanamori, T. & Tachibana, H., 2005. Experimental study on stage acoustics for ensemble performance in chamber music. *Acoustical science and technology*, 26(4), pp.345–352.

Vigeant, M.C., Wang, L.M. & Holger Rindel, J., 2008. Investigations of orchestra auralizations using the multi-channel multi-source auralization technique. *Acta Acustica united with Acustica*, 94(6), pp.866–882.

Vigeant, M.C., Wang, L.M. & Rindel, J.H., 2011. Objective and subjective evaluations of the multi-channel auralization technique as applied to solo instruments. *Applied Acoustics*, 72(6), pp.311–323.

Vorländer, M., 1995. International round robin on room acoustical computer simulations. In 15th International Congress on Acoustics. Trondheim, Norway.

Yadav, M., Miranda Jofre, L.A., Cabrera, D. & Martens, W.L., 2012a. A method for simulating autophony with auralized oral-binaural room impulse responses. Submitted to the 133rd Audio Engineering Society Convention. San Francisco, CA, USA.

Yadav, M., Cabrera, D. & Martens, W.L., 2012b. A system for simulating room acoustical environments for one's own voice. *Applied Acoustics*, 73(4), pp.409–414.

Yadav, M., Cabrera, D. & Martens, W.L., 2011. Auditory room size perceived from a room acoustic simulation with autophonic stimuli. *Acoustics Australia*, 29(3), pp.101-105.

Zmölnig, J., Ritsch, W., & Sontacchi, A. (2003). The IEM-Cube. *Proceedings of the 2003 International Conference on Auditory Display*. Boston, MA.

Zolzer, U. ed., 2011. *DAFX: Digital Audio Effects* 2nd ed, Wiley.

Zotter, F., Pomberg, H., & Noisternig, M. (2010). Ambisonic Decoding With and Without Mode-Matching: A Case Study. *Procedings of the 2nd International Symposium on Ambisonics and Spherical Acoustics*. Paris, France.