# COMPARISON OF SPATIAL AUDIO TECHNIQUES FOR USE IN STAGE ACOUSTIC LABORATORY EXPERIMENTS

*Iain Laird**

Digital Design Studio
Glasgow School of Art, UK
I.Laird1@gsa.ac.uk

*Damian Murphy*

Audio Lab Department of Electronics
University of York, UK
damian.murphy@york.ac.uk

*Paul Chapman*

Digital Design Studio
Glasgow School of Art, UK
p.chapman@gsa.ac.uk

## ABSTRACT

Real-time auralisation systems are increasingly being used by researchers aiming to observe how particular stage and auditorium configurations affect a musician's performance technique. These experiments typically take place in controlled laboratory conditions equipped with auralisation systems capable of reproducing the 3D acoustic conditions of a performance space in response to a performing musician in real-time. This paper compares the performance of First Order Ambisonics and Spatial Impulse Response Rendering in terms of both objective measurements and subjective listening tests. It was found that both techniques spatialised single reflections with similar accuracy when measured at the sweet spot. Informal listening tests found that the techniques produced very similar perceived results both for synthesised impulse responses and for measured stage acoustic impulse responses.

## 1. INTRODUCTION

In order to investigate specific subjective effects of stage acoustic conditions on a performing musician, it is necessary to introduce musician test subjects into known stage environments and allow them to play in the space, noting their subjective reaction to specific objective variables. The most straightforward way of providing this environment is to perform the experiments in existing performance spaces. However, gaining access to performance spaces can be costly and often it is not possible to control specific aspects of the acoustic response. Therefore, the alternative is to develop a laboratory system which is capable of presenting a test subject with controlled acoustic conditions allowing, for example, a musician to play in a virtual version of a performance space.

Such research has been emerging since the 1980's with Gade [1] pioneering an approach using electronic delays and reverberation chambers to create virtual versions of concert halls and investigating specific phenomena experienced by musicians. Digital audio and spatial audio techniques have moved on significantly since then and it is now possible to provide a listener with a much more accurate 3D simulation of a soundfield which can be adjusted to allow certain phenomena to be investigated in more detail.

* This work was supported by Arup Acoustics. Thanks to all the listening test volunteers and the Glasgow City Halls for allowing access to the venue

Recently, a number of authors have made use of real-time convolution [2] and Ambisonic recording/playback techniques [3] in order to emulate stage acoustic conditions for a performing musician in real-time. The use of First Order Ambisonics (FOA) provides a convenient and efficient way of capturing, analysing, transforming and recreating 3D soundfields in laboratory conditions. However, a real-time auralisation system capable of reproducing an accurate and natural sounding acoustic environment continues to present a significant challenge. FOA-based systems, are known to have a limited spatial resolution and due to the highly correlated nature of the loudspeaker signals, can often result in blurred and coloured reproduction of sound sources at the sweetspot [4]. Other techniques, such as Higher order Ambisonics (HOA) [5] have recently been used to increase the spatial accuracy of an auralised soundfield, however the lack of widespread availability of HOA microphones can in some cases prohibit the use of measured impulse responses.

Spatial Impulse Response Rendering (SIRR) [4], is a more recently developed spatial audio technique which is capable of providing detailed directional analysis, complex modification and reproduction of 3D impulse responses over arbitrary loudspeaker arrays. It is a perceptually motivated approach which analyses a 3D impulse response for physical properties that will transform into human auditory localisation cues. It then synthesises appropriate loudspeaker feeds aiming to recreate a natural sounding soundfield with the equivalent spatial impression. SIRR provides an attractive alternative to FOA-based real-time auralisation systems as it is possible to manipulate and analyse elements of an impulse response in much finer detail.

This paper presents a comparison of SIRR and FOA-based real-time auralisation systems for use in the context of stage acoustic laboratory tests. A series of objective and subjective tests were performed to indicate which technique is more appropriate for these types of experiments. The tests aimed to objectively compare the spatialisation quality of auralised impulse responses using each technique by measuring at the sweetspot of an auralisation system. Subjective tests were also carried out to determine if there were any audible differences using each technique when auralising various target acoustic conditions. If a SIRR-based system was demonstrated to produce similar or better objective and subjective results than an FOA-based system then it would be an initial indication that using SIRR is a viable option for stage acoustic experiments and therefore the various advantages it has can be exploited in the future.

The structure of the paper is as follows: The first section begins by briefly describing the typical architecture of a real-time auralisation system used in stage acoustic laboratory tests. It will then describe the operation of SIRR for both analysis and re-synthesis of 3D impulse responses and compare the operating performance against FOA based systems using a series of objective tests. Finally, the paper will report on an informal listening test which aimed to ascertain if naive listeners could detect any subjective differences between 3D soundfields recreated using SIRR and FOA.

## 2. REAL-TIME AURALISATION SYSTEMS

A real-time auralisation system measures the direct sound created by a musician which is then processed by a computer capable of performing real-time convolution of the direct sound with a 3D impulse response of a performance space. In FOA-based systems, the direct sound is convolved with the four channels of a B-format impulse response representing the target space. The processed audio is passed to a decoder matrix which produces a number of speaker feeds which play back the auralised sound of the performance space back to the musician over a loudspeaker array. Figure 1 shows the layout of a typical FOA-based real-time auralisation system. A SIRR-based system operates in a similar way with the exception that the impulse response has been decomposed into a number of impulse responses (one per loudspeaker) and so there is no need for a decoder stage.

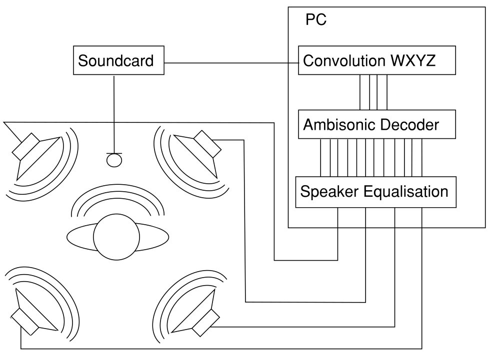



Figure 1: *System diagram of a typical FOA-based real-time auralisation system (only 4 loudspeakers are shown for clarity)*

A real-time auralisation system similar to that shown in figure 1 was constructed in the Arup-DDS SoundLab situated in Glasgow, UK. The SoundLab is an acoustically controlled space with dimensions 4m (l) x 6m (w) x 2.5m (h). The average T20 of the space at 500Hz is approximately 0.15s. The SoundLab has a measured background sound pressure level ($L_{AF90}$) of approximately 18dBA. The SoundLab features a 16-channel 3D loudspeaker array comprising of three rings of Yamaha MSP5A active loudspeakers arranged in a 4 - 8 - 4 arrangement as shown in figure 4. The speaker system was equalised to ensure a flat frequency response and an equal level contribution at the sweetspot.

For the purposes of stage acoustic auralisation it is neither necessary or practical to auralise the direct sound or the floor reflection as they are produced by the instrument and solid floor respectively. By silencing any audio before the first required reflection and then truncating by a value equal to the system latency, the effective latency of the system can be minimised. Typically, the direct sound is measured using directional microphones positioned close to the musician in order to minimise unwanted feedback. In this particular system the direct sound is measured using a single directional microphone, however further improvements could be made by using multiple microphones positioned around the musician to ensure the complex, time-varying radiation characteristics of the instrument are captured.

## 3. SPATIAL IMPULSE RESPONSE RENDERING

Spatial Impulse Response Rendering (SIRR) is a spatial audio technique which allows a 3D soundfield to be rendered to an arbitrary speaker layout [4]. The technique involves the analysis of a soundfield to obtain its directional properties and subsequent synthesis of the resulting diffuse and non-diffuse cues to recreate a perceptually equivalent soundfield. A 3D soundfield can be measured with an Ambisonic microphone and analysed in the time-frequency domain to produce a pressure signal with accompanying meta-data carrying information regarding the direction of arrival and diffuseness of each time-frequency element. The meta-data is then used to reconstruct the soundfield using amplitude panning techniques. SIRR has found many different applications including soundfield analysis [6], high quality auralisation of room acoustics [4] and parametric spatial audio effects [7].

### 3.1. Analysis

The direction of arrival of a sound can be estimated by obtaining the active sound intensity which is a product of the sound pressure $p(t)$ and particle velocity vector $u(t)$. This describes the transfer of energy of the soundfield and therefore the opposing vector will describe the direction of arrival of the sound. This can be achieved by analysing the B-format output of an Ambisonic microphone where the omnidirectional signal, $W(t)$, is assumed to be proportional to the pressure $p(t)$. The remaining orthogonal, figure-of-eight pressure-gradient signals $X(t)$, $Y(t)$ and $Z(t)$ can be considered as proportional to the components of the particle velocity $u(t)$. Therefore the active intensity can be obtained using equation (1) and the direction of arrival found using equation (2) and (3) giving azimuth and elevation respectively. Where $\acute{X}(\omega) = (X(\omega)e_x + Y(\omega)e_y + Z(\omega)e_z)$, "$*$" denotes complex conjugation and $Z_0 = \rho_0 c$ is the acoustic impedance of air.

$$I_\alpha(\omega) = \frac{\sqrt{2}}{Z_0}\Re\{W^*(\omega)\acute{X}(\omega)\} \quad (1)$$

$$\theta(\omega) = \tan^{-1}\left[\frac{-I_y(\omega)}{-I_x(\omega)}\right] \quad (2)$$

$$\phi(\omega) = \tan^{-1}\left[\frac{-I_z(\omega)}{\sqrt{(I_x^2(\omega) + I_y^2(\omega))}}\right] \quad (3)$$

Diffuseness can be estimated by obtaining the proportion of sound energy contributing to the net transport of energy and can be calculated using equation (4). This produces a value between 1 and 0 for each time-frequency element characterising the sound as either diffuse or non-diffuse. By multiplying the audio signals by $\sqrt{\psi}$ or by $\sqrt{1-\psi}$ the audio signals can be separated into diffuse and non-diffuse signals respectively and re-synthesised with appropriate spatial audio techniques as described below.

$$\psi(\omega) = 1 - \frac{2Z_0 \|\Re\{W^*(\omega)\acute{X}(\omega)\|}{|W(\omega)|^2 + |\acute{X}(\omega)|^2/2} \tag{4}$$

### 3.2. Synthesis

Synthesis of the audio signals takes place in the frequency domain by transforming the audio signals using the Short Time Fourier Transform (STFT) and applying the meta-data obtained in the analysis to the audio signals before using the Inverse Short Time Fourier Transform (ISTFT) to produce the output time-domain audio signals. The analysis-resynthesis process is shown in Figure 2.

Non-diffuse sound sound synthesis aims to reproduce coherent reflections as point-like sources and is typically implemented using Vector Base Amplitude Panning (VBAP) [8]. VBAP is an amplitude panning technique that allows sounds to be panned around a periphonic loudspeaker array using vector calculation to determine the level of a local triplet of loudspeakers. The diffuse sound synthesis aims to recreate the reduced interaural coherence produced by the diffuse sound energy. This is achieved by decorrelating the sounds identified as being diffuse and distributing equally to each loudspeaker. Decorrelation can can be implemented in a number of different ways however in this study decorrelation via time-varying phase randomisation was used which has been reported to give acceptable results when re-synthesising impulse responses [9].

High quality implementations of SIRR makes use of all B-format signals, applying the meta-data to a set of decoded signals for each loudspeaker using virtual microphone principles. This has been observed to provide better directional separation, natural decorrelation and overall higher audio quality. The directivity factor has been found to produce favourable results when set as a dipole microphone pattern and angled towards each loudspeaker [10]. It has also been shown that the use of virtual microphones can affect the correct reproduction of energy in both diffuse and non-diffuse sound. It has been demonstrated that it is possible to apply correction gains to ensure the correct ratio of diffuse and non-diffuse components [10, 9].

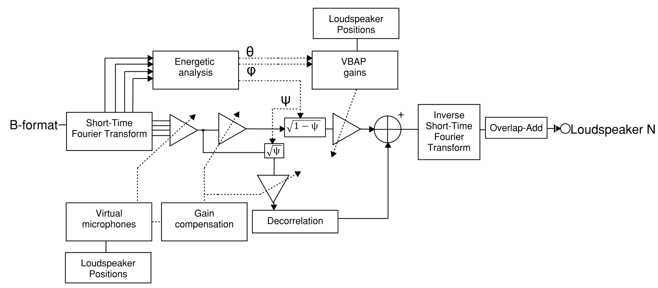



Figure 2: *System diagram of synthesis technique shown for a single loudspeaker channel*

The synthesis phase of SIRR produces a time-domain impulse response signal per loudspeaker channel which can then be utilised by a multichannel convolving reverberator to recreate the acoustic conditions of the target space at the sweetspot.

## 4. OBJECTIVE ANALYSIS

The intensity vectors obtained in the analysis phase (as described in section 3.1) can also be used to evaluate characteristics of 3D impulse responses. When a coherent reflection or plane wave is analysed using SIRR, the intensity vectors all tend to point in the direction of arrival at the time of arrival. Conversely, in diffuse (or reverberant) conditions, the intensity vectors are distributed in a more stochastic fashion. Conditions measured on stage typically consist of a number of coherent early reflections followed by a diffuse reverberant decay. Therefore, observing the angular distribution of intensity vectors in an impulse response will allow the presence of reflections to be detected [6]. Furthermore, the mean direction of the intensity vectors will indicate the direction of arrival of each reflection.

Spherical variance, $\sigma$ (5), can be used to assess the angular distribution of the intensity vectors for each time step. It is defined as the magnitude of the mean resultant vector, $S(t)$ (6) where $\|\cdot\|$ determines the magnitude of the enclosed vector. The intensity vectors $I_i(t)$ are normalised prior to calculation of $S(t)$ to provide more robust results as demonstrated in[6].

$$\sigma(t) = 1 - \|S(t)\| \tag{5}$$

$$S(t) = \frac{1}{X}\sum_{i}^{X} I_i(t) \tag{6}$$

Where $I_i(t)$ is the $i^{th}$ frequency band for time step $t$ and $X$ is the total number of frequency bins used in the analysis [6, 11]. The mean angular direction for each time frame can be subsequently computed using equations (2) and (3). For stage acoustic impulse responses, $\sigma$ will be close to zero when $t = 0$ (as the direct sound arrives) and will increase quickly to a maximum value less than one as the impulse response becomes increasingly diffuse with time. The arrival of a coherent reflection will produce a localised trough in this response with a magnitude dependent on the nature of the reflection.

When using a short time window and low hop size in the time-frequency analysis, a single reflection may be identified a number of times with slightly different results per iteration resulting in a number of points representing the direction of arrival per reflection. A Gaussian Mixture Model (GMM) can be used to identify clusters of these points when arranged by diffuseness and direction of arrival. The component-mean (centre) of the cluster can then be used to estimate a direction of arrival.

## 5. TEST METHODOLOGY

In order to compare the performance of each spatialisation technique in the context of stage acoustic laboratory experiments, a real-time auralisation was set up in the SoundLab using known impulse responses rendered using FOA or SIRR. The impulse response of the auralised space within the SoundLab was measured at the sweetspot in order to emulate a musician using the space. This was achieved by positioning a directional loudspeaker and ambisonic microphone in the sweet spot of the loudspeaker array to represent the musician's instrument and head respectively.

The impulse responses used in the test consisted of either a single synthesised reflection arriving from a single direction or a measured stage impulse response obtained during a survey of the Grand Hall, Glasgow City Halls. The stage impulse response was obtained by measuring in the venue using a Genelec 1029A Active loudspeaker and Soundfield ST350 Ambisonic microphone arranged in a manner emulating the instrument and head of a musician respectively. The apparatus was positioned down-stage right approximately 4m away from a nearby side wall. The average mid-frequency reverberation time (T30, 500Hz) at this position was found to be approximately 1.75 seconds. The synthesised impulse response consisted of a single, non-diffuse reflection with a time delay of 60ms relative to the direct sound which was panned using ambisonic panning techniques to various angles of azimuth ($0°$, $\pm 60° \pm 90°$, $\pm$ $110°$). The amplitude of this reflection was set to be -6dB below the direct sound.

Both the measured and synthesised FOA impulse responses were processed as described in section 3 to obtain an impulse response for each loudspeaker in the SoundLab. The analysis and re-synthesis was implemented in the time-frequency domain using a MATLAB script. An STFT was used with a 16-sample Hanning window (with 16 samples of zero padding) and a hopsize of 4 samples. Signal reconstruction utilised an ISTFT alongside the OverLap-Add method (OLA) in order to obtain as close to perfect reconstruction as possible. The window size was chosen to ensure that the highly transient nature of the impulse response signals were preserved. It is noted however that the small window size will require a compromise in terms of frequency resolution due to the inherent trade-off between time and frequency resolution of the STFT.

A Genelec 1029A Active loudspeaker was mounted on a tripod at a height of 100cm above the floor (height to top of main driver) and a radial distance of approximately 50cm from the receiver to the centre of the loudspeaker. The receiver was an Soundfield ST350 Ambisonic microphone which was located in the sweet-spot at a height of 130cm shown in Figure 3. Impulse responses were measured using a 10-second, logarithmically swept sinusoidal signal generated in MATLAB. The sine sweep was played through the loudspeaker and measured simultaneously at a sampling frequency of 44.1kHz and a 32-bit floating-point bit depth. Impulse responses were extracted from the recorded sine sweeps by convolution with the inverse of the original sweep as demonstrated by Farina [12].

The direct sound from the loudspeaker was measured using an M-Audio Luna cardioid condenser microphone positioned on axis between low and high frequency drivers of the loudspeaker at a 10cm distance. As with previous experiments [2], the direct sound in this experiment was measured using a single directional microphone. This is a noted simplification as a musical instrument exhibits complex and time-varying radiation characteristics which cannot be measured using a single microphone. The sound source used to measure the stage impulse responses is identical to that used in this experiment therefore the error in using a single microphone to measure the direct sound is minimised. It will be necessary for future stage acoustic laboratory tests to consider the impact of this.

The direct sound was input into the auralisation system as described in section 2. When auralising using FOA, real-time convolution was performed in Max MSP using two, 4 SIR2 VST convolution engines, the resultant convolved audio was decoded to the 16-channel loudspeaker array in the SoundLab using a Gerzonic Decopro ambisonic decoder set to a "Max-RE" type for all frequencies. When auralising using SIRR, the number of convolution objects increased to 16 (one per loudspeaker) and the ambisonic decoder is removed.

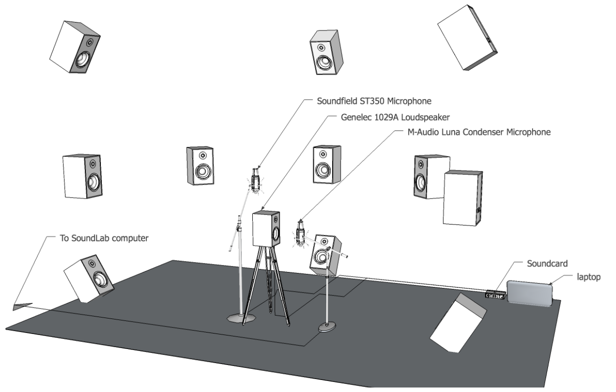



Figure 3: *Model of experimental set up in SoundLab (loudspeakers on one side have been omitted for clarity)*

## 6. OBJECTIVE RESULTS

Figure 4 shows an example set of objective results for an auralised reflection arriving at an angle of $-90°$ (anticlockwise), 60ms after the direct sound at a level of -6dB below the direct sound. The omnidirectional amplitude response of this scenario is shown for both FOA and SIRR. The direct sound occurs shortly after $t = 0$ and the reflection clearly arrives at $t = 60ms$. The decay seen after each of these events is caused by the SoundLab acoustic response. Overlaid on this plot is the spherical variance for both techniques. It can be seen the spherical variance is very similar for the first 50ms which starts at a value of zero when $t = 0$ and increases rapidly during the SoundLab acoustic decay. After $t = 50ms$, there are clear differences for each technique when the reflection arrives. It can be seen there is a sharp reduction in spherical variance when the reflection arrives which is of greater amplitude when using SIRR than FOA.

Figures 5(a) and 5(b) show the angle of the mean resultant vector for the first 0.2 seconds of the synthesised impulse response using FOA and SIRR respectively. For clarity, the plots show the analysis for parts of the impulse response that are below a mean diffuseness value of 0.55. The smaller dots show the analysed angle of arrival while the larger dots show the GMM estimate. The expected angle of arrival is shown by a cross which in this case is positioned at $-90°$.

When using the GMM estimation method to estimate the angle of arrival from clusters of points, errors can be introduced into the estimation as the location of a component-mean is influenced by nearby clusters. These clusters can be created by periods of silence or other nearby reflections. In this case, the adjacent clusters are caused by the room acoustic response of the SoundLab. Therefore for each GMM estimation, three component-means were calculated. This is to ensure that the auralised reflection is estimated by one component-mean whereas the response of the SoundLab is estimated separately, the results of which are then discarded.

It can be seen in both cases (Figures 5(a) and 5(b)) that one component-mean is very close to the expected angle of arrival while the remaining estimates are related to the SoundLab acoustic response.

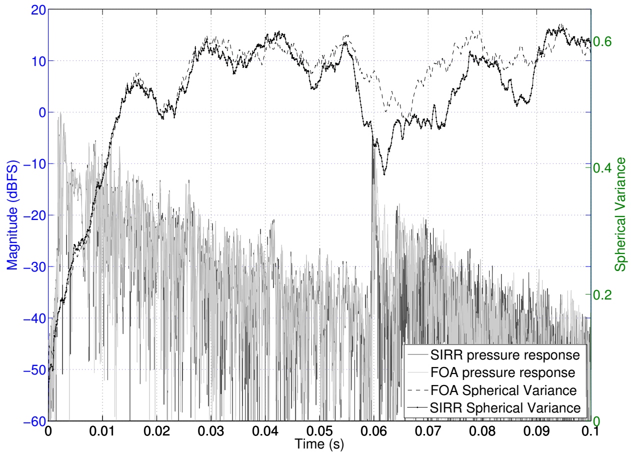
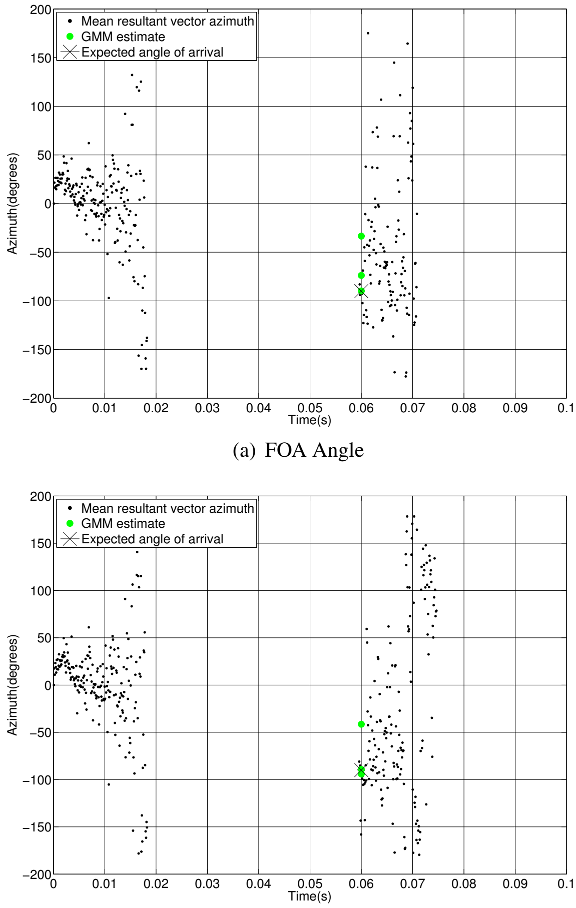



Figure 4: *Composite plot showing measured FOA and SIRR Sound Pressure Level envelopes (dBFS) and the associated spherical variance measured for each technique. This example shows the direct sound from the loudspeaker (including the early response of the SoundLab) and a single reflection occurring at $t = 60ms$.*




(a) FOA Angle

(b) SIRR angle

Figure 5: *Plots of the direction of the mean resultant vector over time for an impulse response containing a reflection at $t = 60ms$. Results with diffuseness $> 0.55$ have been omitted for clarity. The cross represents the expected angle of arrival ($-90°$). The larger dots (GMM component means) show the estimated angle of arrival*

## 7. LISTENING TESTS

Informal listening tests were undertaken in the SoundLab to evaluate the perceived performance of an FOA and SIRR based systems. In previous studies [1] it was found that because the musician was generating a sound that was subsequently auralised, it is highly unlikely the musician would be able to exactly repeat that sound affecting the reliability of experiments. In order to specifically compare the performance of the auralisation system it was necessary to design a repeatable experiment where the source signals remained unchanged with each repetition. Therefore, this listening test was designed as a passive test where test subjects were asked to sit in the sweet-spot of the loudspeaker array and listen to pairs of brief musical samples which were played through a directional loudspeaker positioned in front of them (imitating the musical instrument) and auralised in real-time with known impulse responses using FOA and SIRR. In this experiment, the test subject is not engaged in the physical act of playing their instrument and so is able to focus their attention solely on the audio stimuli presented to them. They will therefore be more sensitive than performing musicians to subtle differences in the presented acoustic response. Therefore this test should be viewed only as conservative comparison between the two spatialisation techniques.

The listening test was an A/B (hidden reference) test where listeners compared pairs of musical samples (where each sample had been auralised with an impulse response then rendered using FOA or SIRR) and asked to rate on a five-point Likert scale how similar or different they thought the two sounds were (1 being most similar and 5 being most dissimilar). The test pairs were short musical samples played through the loudspeaker in front of the test subject (a short cello sample playing legato (Source 1), short clarinet sample playing staccato (Source 2) and a sustained long note from a clarinet (Source 3)). This sound was auralised in real-time with an impulse response consisting of either (a) - a synthesised coherent reflection or (b) - a measured stage impulse response obtained in the Grand Hall of Glasgow City Halls or (c) - no synthesised acoustic response. The test subjects could listen to each excerpt as many times as they liked before recording their answer. The test subjects were not given any visual reference and were asked to face forward at all times. A number of null tests were introduced where both samples were the same in each pair and a single example test was presented to the listener at the beginning which was not included in the results.

There were 47 randomised combinations of stimuli in total. There were six volunteers, all between the ages of 24 and 32 (4 male, 2 female). Most of the test subjects had a background in audio engineering or acoustics, all of the volunteers had some experience of music performance. All test subjects reported no significant hearing loss.

It was expected due to the different rendering methods used, the test subject would consistently identify one method over another due to the differing ability to accurately place and reproduce the reflection. Consequently, when the test subject was asked to rate the similarity of (a) no reflection against (b) SIRR or FOA rendered reflection, it was expected that they would report a larger difference with one technique than the other. Furthermore, it was expected that the test subjects would report significant audible differences when a musical sample, auralised with a stage acoustic impulse response, was rendered using (a) SIRR or (b) FOA techniques.

## 8. LISTENING TEST RESULTS

The results in figure 6 show how similar or dissimilar the participants thought the musical pairs were when one of the samples was auralised with a single reflection using FOA or SIRR and the other sample was played without a simulated reflection (i.e. direct sound only). It can be seen that in most cases, auralising with either technique produces similar median scores throughout. Furthermore, there is no discernible pattern by which one technique results in larger differences than the other.

For each reflection angle, it can be seen from the responses that the staccato clarinet sound source (source 2) resulted in reflections being identified more easily than the clarinet tone or legato cello. There is also a general indication that the presence of a reflection was more easily detected when there was a high angular separation between the reflection and the sound source.

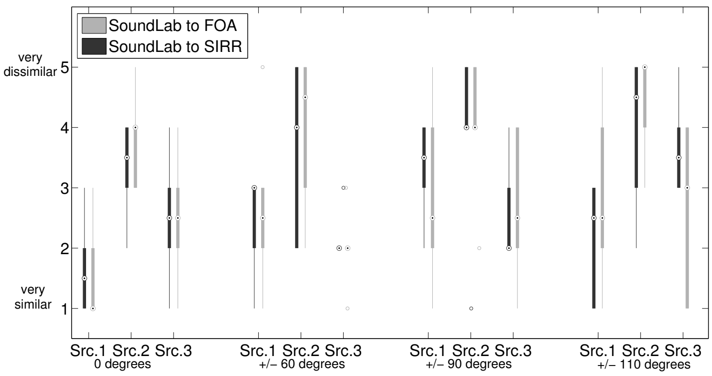



Figure 6: *Listening test results for sounds auralised with a single reflection rendered with FOA or SIRR compared to no reflection. Thick lines indicate 25 and 75 percentiles, thinner lines show the extremities of the data points, dots within boxes indicate the median while circles indicate outliers*

The results in figure 7 show how similar or dissimilar the participants thought the musical pairs were when musical samples were auralised with a stage acoustic impulse response rendered with FOA or SIRR. The results show that test subjects could discern a slight difference between the musical samples and that this difference was consistent even if the source type was altered.

## 9. DISCUSSION

Figure 4 showed that for a single auralised reflection, the spherical variance is lower when it is rendered using SIRR than when using FOA. This illustrates that the intensity vectors, indicating the direction of arrival of the reflection, are more tightly clustered

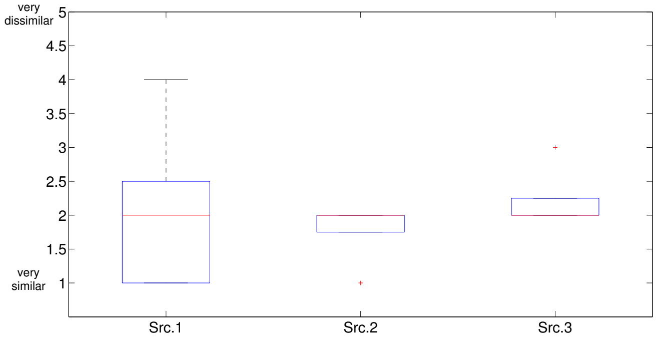



Figure 7: *Listening test results for sounds auralised with a measured impulse response rendered with FOA compared with SIRR. Central lines indicate the median response while box edges indicate 25th and 75th percentiles, outliers are indicated by crosses*

when using SIRR and therefore less ambiguous in terms of spatialisation. This is to be expected for the example shown as the reflection arrived from the same direction as one of the loudspeakers. When using SIRR, this reflection would be rendered exclusively with VBAP (due to it being non-diffuse) and therefore only one loudspeaker would be active when recreating this sound. In FOA, however all the loudspeakers are active at all times, therefore more loudspeakers would be contributing to recreate this reflection which may have contributed to a wider distribution of intensity vectors and hence a higher value of spherical variance. When single reflections were auralised from a direction between two loudspeakers, the spherical variance increased slightly when using SIRR but remained mostly unchanged when using FOA. This is consistent with operation of both spatial audio techniques where the localisation quality of FOA can be made to be largely independent of the direction of arrival. The localisation quality of VBAP will reduce slightly when the direction of arrival is between two loudspeakers separated by a large distance.

The GMM was shown to perform well when estimating the angle of arrival of reflections. It was found however that the SoundLab acoustic response introduced a noise floor into the estimation which increased as the time of arrival of the reflection decreased. This method is useful for assessing measured impulse responses but requires improvement for use in assessing auralised reflections in non-anechoic conditions. In this case, a more rigorous clustering approach could improve the accuracy of the results.

The results presented in figure 6 and figure 7 provide an initial indication that for auralisation of single reflections and measured stage acoustic impulse responses, SIRR and FOA perform equally well in the context of stage acoustic experiments although subtle differences can be discerned between the two techniques. A general trend in the listening test results indicate that in the presence of a directional sound source (representing a musical instrument), a listener may have more difficulty detecting the presence of a reflection when it arrives at a similar angle to the sound source. This is thought to be predominantly due the masking effects caused by the musical instrument. Similarly, the results also imply that detection of reflections can be affected by the type of musical expression. In this experiment, the reflections appear to have been detected more easily when the sound source was a clarinet playing staccato notes. This is thought to be due to the increased transients, coupled with note spacing allowing the auralised reflection to be more easily detected. This however does not appear to be the case when the sound is auralised using a stage acoustic impulse response as the median results in figure 7 are very similar.

Overall, the results provide an initial indication that SIRR and FOA-based auralisation techniques perform equally well in the context of stage acoustic laboratory tests for a listener positioned in the sweetspot. However, mainly due to the low number of untrained participants there is significant uncertainty within the results, therefore further study with musician test subjects is required in order to allow for a more concrete comparison. Considering the potential advantages that SIRR offers in terms of detailed modification and analysis of impulse responses, this study provides an indication that SIRR is a viable technique for future research. This could potentially allow researchers to take advantage of the complex transformations and analysis that SIRR is capable of in addition to providing a simulation environment where the listeners impression of the soundfield is less dependent on them being located exactly in the sweetspot.

## 10. CONCLUSIONS

This article has demonstrated how SIRR can be used in the context of stage acoustic laboratory experiments to provide a detailed analysis of an impulse response presented to a test subject and also to re-synthesise a measured B-format impulse response using a combination of VBAP and de-correlated speaker feeds. Informal listening tests provided an initial indication that passive listeners could detect small differences between impulse responses auralised in real-time with SIRR and with FOA. Future work will focus on development of the SIRR technique for use specifically in stage acoustic laboratory experiments.

## 11. REFERENCES

[1] A. C. Gade, "Subjective room acoustic experiments with musicians," ISSN 0105-3027 32, Technical University of Denmark, 1982.

[2] K. Ueno, T. Kanamori, and H. Tachibana, "Experimental study on stage acoustics for ensemble performance in chamber music.," Acoust. Sci. Tech. 26, 4 (2005), 2005.

[3] J. Brereton, D. Murphy, and D. Howard, "A loudspeaker-based room acoustics simulation for real-time musical performance," Proc. of the 25th UK Audio Engineering Society conference in association with the 4th International Symposium on Ambisonics and Spherical Acoustics, 2012.

[4] J. Merimaa and V. Pulkki, "Spatial impulse response rendering," Naples, Italy, October 2004, Proceedings of the 7th International Conference on Digital Audio Effects (DAFx), pp. 139–144.

[5] A. Guthrie, S. Clapp, and N. Xiang, "Using ambisonics for stage acoustics research," Toronto, Canada, 2013, Proc. of the International Symposium on Room Acoustics.

[6] S. Tervo, T. Korhonen, and T. Lokki, "Estimation of reflections from impulse responses," Melbourne, Australia, August 2010, Proc. of the International Symposium on Room Acoustics.

[7] A. Politis, T. Pihlajamaki, and V. Pulkki, "Parametric spatial audio effects," York, UK, September 2012, Proc. of the 15th International Conference on Digital Audio Effects (DAFx-12).

[8] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," Proc. of the Audio Engineering Society, June 1997, vol. 45.

[9] T. Pihlajamaki, "Multi-resolution short-time fourier transform implementation of directional audio coding," M.S. thesis, Helsinki University of Technology, Faculty of Electronics, Communications and Automation, Department of Signal Processing and Acoustics, Espoo, Helsinki, August 2009.

[10] J. Vilkamo, "Spatial sound reproduction with frequency band processing of b-format audio signals," M.S. thesis, Helsinki University of Technology, Faculty of Electronics, Communications and Automation, Department of Signal Processing and Acoustics, Espoo, Helsinki, May 2008.

[11] A. Southern and L. Savioja, "Spatial high frequency extrapolation method for room acoustic auralization," York, UK, September 2012, Proc. of the 15th International Conference on Digital Audio Effects (DAFx-12).

[12] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," Paris, France, 2000, Proceedings of the 108th Audio Engineering Society Conference.