# Acoustic Source Localisation In An Urban Environment Using Early Reflection Information

Francis Stevens and Damian T. Murphy
Audio Lab, Department of Electronics, University of York, Heslington, YO10 5DD, York, UK.

**Abstract**

Acoustic source localisation is the use of recorded information to determine the point of origin of a given sound. It has applications in military threat detection, forensics, and the study of environmental acoustics. Impulse responses recorded in a semi-enclosed urban environment have shown early reflections to be the dominant acoustic feature, with the majority of directional information present in the horizontal plane. This paper presents a source localisation algorithm that uses this information. Spatial Impulse Response Rendering (SIRR) analysis is used to extract reflection information from B-format impulse response measurements. Reverse ray-tracing is then used with a 2D geometric representation of the environment to estimate the source position. When used for recordings made in an enclosed and highly reverberant environment, the localisation performance suffers due to the lack of highly distinct early reflections.

**PACS no. 43.55.Cs, 43.60.Jn**

## 1. Introduction

The study of acoustic propagation in indoor spaces for auralisation is a well established discipline. However, the acoustic properties of outdoor environments are not well studied. In order to develop this understanding an acoustic survey was conducted at the University of York to generate a set of spatial (B-format) impulse responses (IRs) recorded in a semi-enclosed urban environment [1]. Analysis of the survey results indicated early reflections to be the dominant acoustic feature, with the open nature of the space limiting the most significant acoustic information present to the horizontal plane.

These results led to the development of a localisation algorithm that makes use of this information and a 2D model of the recording space. This paper presents the design and implementation of this algorithm, starting with a brief review of the conducted acoustic survey, followed by the main elements of the algorithm design: early reflection information extraction and ray tracing.

Results are then presented, including evaluation of changing localisation performance with multiple variables, including: the sound source used, the distance between source and receiver, the presence of line-of-sight between source and receiver, and the number of recordings used in the localisation. The paper concludes with the application of this method to an enclosed, indoor space, with a highly reverberant acoustic with respect to early reflections.



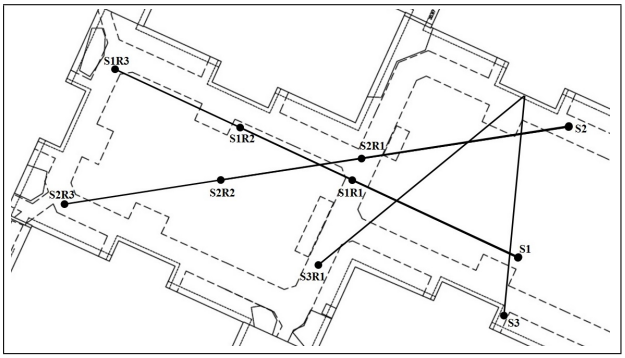


Figure 1. Plan view of the measurement site, with source and receiver positions, and the direct sound path in each case indicated.

## 2. Recorded Impulse Responses

The IRs used in the development and testing of the localisation algorithm were recorded in a semi-enclosed residential courtyard on the University of York campus, York, UK. This site (shown in Figure 1) was chosen for the recording work for several reasons: a 2.4m boundary wall covering all but the south east entrance gives some acoustic isolation from surrounding structures; the size of the courtyard (c. 70m by 30m) and open space between the buildings makes it suitable for examining the acoustic behaviour between multiple buildings, and small enough to easily identify the effects of this behaviour; the buildings are comprised of mainly planar, largely orthogonal, surfaces making the site easy to model and reflection paths easy to trace.

Seven receiver positions were chosen for three source positions (indicated in Figure 1). This included two axes formed of three sets of receivers: one approximating the diagonal span of the courtyard (positions S1R1-S1R3), and the other along the length of the courtyard; parallel to the frontage of the buildings (positions S2R1-S2R3). The final source-receiver pair (recording position S3R1) was positioned to break line-of-sight between source and receiver.

IRs were recorded at each position using a starter pistol and an exponential sine sweep. A SoundField ST450 B-format microphone was used for recording each IR, and a Genelec S30D was used to reproduce the sine sweeps (frequency range 22 Hz - 22 kHz, duration 60 seconds). A full description of the recording work can be found here [1], and a full set of IRs is available online on OpenAIR at [2].

### 2.1. SIRR Analysis

The directional information encoded in B-format recordings can be expressed by calculation of the instantaneous intensity vector, **I**. The B-format signal is divided into discrete time frames, with a short-time Fourier transform (STFT) performed on each channel. The resultant frequency domain signals can be used to estimate the intensity vector:

$$\mathbf{I}(\omega) = \frac{\sqrt{2}}{Z_0}\mathscr{R}\{W^*(\omega)\mathbf{U}(\omega)\} \tag{1}$$

where $\mathbf{U}(\omega)$ is the vector $[X(\omega), Y(\omega), Z(\omega)]$, $Z_0$ is the characteristic acoustic impedance of the air, and * denotes the complex conjugate [3]. A time-frequency distribution of **I** vectors can be overlaid on a spectrogram of the recording's omni-directional (W-channel) response to generate quiver plots (such as the one shown in Figure 2) allowing for the concurrent analysis of the magnitude and direction of arriving acoustic energy. Calculation of **I** is one of the steps involved in Spatial Impulse Response Rendering (SIRR), a method of rendering spatially captured IRs over a multichannel loudspeaker system [4].

## 3. Algorithm Design

The proposed algorithm aims to use the information presented in SIRR analysis quiver plots, a 2D model of the recording space (i.e. a 'slice' in the horizontal plane), and knowledge of the position and orientation of the recording microphone within the space, to trace ray paths representing identified early reflections in the space. Where these ray paths cross will identify the source position. This section will cover each step of the algorithm design: extracting a set of early reflection information from the recorded IRs, tracing identified reflection paths, and estimating the source position from the plotted paths.

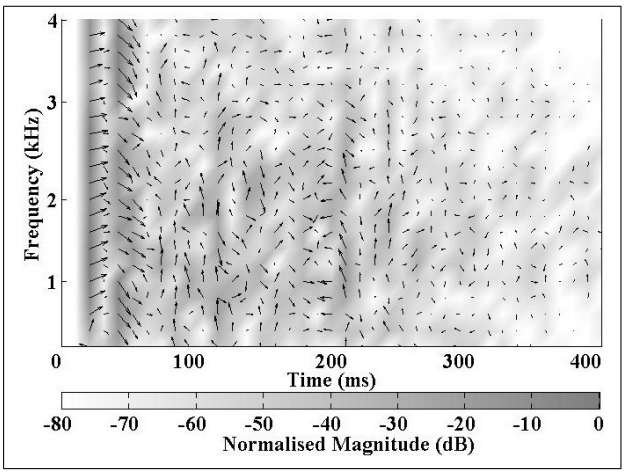


Figure 2. An example quiver plot showing results from SIRR analysis of the horizontal plane of a recording made at position S1R1.

### 3.1. Early Reflection Information Extraction

To generate a set of early reflection data suitable for ray tracing, the information represented by the quiver plots, such as the one in Figure 2, needs to be converted into a list identifying several properties for each reflection path: magnitude, path length, and direction.

Assuming that each reflection occurs in a single SIRR analysis frame only, the local energy in each frame can be calculated and compared to surrounding frames, indicating where distinct early reflections are present. This local energy calculation can be expressed in discrete time and frequency form as:

$$X_{[T]} = \frac{\sum_{F=f_{min}}^{f_{max}} \mathbf{I}_{[T][F]}}{\left(\sum_{F=f_{min}}^{f_{max}} \mathbf{I}_{[T-1][F]} + \sum_{F=f_{min}}^{f_{max}} \mathbf{I}_{[T+1][F]}\right) \Big/ 2} \tag{2}$$

where the local energy at time $T$, $X_{[T]}$, is the sum of SIRR analysis magnitude results **I** over the frequency range defined from frequency bin $f_{min}$ to $f_{max}$ in time frame $T$, divided by the mean of the same sum performed in the time frames immediately before $(T-1)$ and after $(T+1)$ [5].

Having calculated the local energy in each time frame, a peak detection algorithm is applied that first identifies all of the peaks and valleys in the local energy data for the duration of the IR:

$$X_{[T_P]} = X_{[T-1]} \leqslant X_{[T]} \geqslant X_{[T+1]} \tag{3}$$

$$X_{[T_V]} = X_{[T-1]} \geqslant X_{[T]} \leqslant X_{[T+1]} \tag{4}$$

where $X_{[T_P]}$ and $X_{[T_V]}$ are the identified peak and valley positions respectively, $X_{[T]}$ is the current time frame's local energy, and $X_{[T-1]}$ and $X_{[T+1]}$ are the local energy either side [6]. Once all of the peaks and valleys have been identified, each peak is then assigned a value of

'peakiness' by comparing its magnitude relative to the nearest valleys either side. This is expressed as:

$$H(X_{[T_P]}) = \frac{(X_{[T_P]} - X_{[T_{V-}]}) + (X_{[T_P]} - X_{[T_{V+}]})}{2} \quad (5)$$

where $H(X_{[T_P]})$ is the 'peakiness' of peak $X_{[T_P]}$, and $X_{[T_{V-}]}$ and $X_{[T_{V+}]}$ are the closest valleys either side of $X_{[T_P]}$. Any peak in the data with a calculated peakiness greater than the mean across all values is identified as a time frame in which there is an early reflection.

The direction associated with each identified early reflection is calculated from the circular mean of all of the directions of arriving acoustic energy values in the specified time frame:

$$\hat{\theta} = \arg\left\{\sum_{i=1}^{N} e^{j\theta_i}\right\} \quad (6)$$

where $\hat{\theta}$ is the circular mean of a set of $N$ angles $\theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$, here defined by the **I** vectors in the relevant time frame [7].

The direct sound path length is currently implemented as a user-selected variable, where the user is required to make an initial estimation of direct path length. The length of each subsequent ray path is then calculated using this direct path length, relative time-of-arrival for each path, and an estimation of the speed of sound.

## 3.2. Ray Tracing

A typical stochastic ray tracer algorithm places a source and receiver at known positions in a modelled environment, and 'fires' a large set of evenly distributed rays from the source to evaluate the environment for valid paths, via boundary reflections, to the receiver. A simplified diagram of this process is shown in Figure 3(a).

The localisation algorithm differs from this process as a set of reflection paths with known directions are traced from a known receiver position in order to estimate the source position. This process is represented in plot Figure 3(b).

In order to plot reflections from boundary walls, points of intersection between the plotted paths and the model boundaries must be calculated. The intersection point of two lines in two dimensional space, defined by the coordinates $(x_{11}, y_{11})$ and $(x_{12}, y_{12})$, and $(x_{21}, y_{21})$ and $(x_{21}, y_{21})$ respectively, is given by:

$$x_{int} = \frac{(x_{11}y_{12} - y_{11}x_{12})(x_{21} - x_{22}) - (x_{11} - x_{12})(x_{21}y_{22} - y_{21}x_{22})}{(x_{11} - x_{12})(y_{21} - y_{22}) - (y_{11} - y_{12})(x_{21} - x_{22})} \quad (7)$$

$$y_{int} = \frac{(x_{11}y_{12} - y_{11}x_{12})(y_{21} - y_{22}) - (y_{11} - y_{12})(x_{21}y_{22} - y_{21}x_{22})}{(x_{11} - x_{12})(y_{21} - y_{22}) - (y_{11} - y_{12})(x_{21} - x_{22})} \quad (8)$$

where $x_{int}$ and $y_{int}$ are the $x$ and $y$ coordinates of the intersection point. Once an intersection point has been calculated, the distance between the intersection point and the origin of the incoming ray is subtracted from the total ray path, calculating the path length of the outgoing ray. This process is repeated until the ray's entire length has been plotted or the maximum reflection order is exceeded.

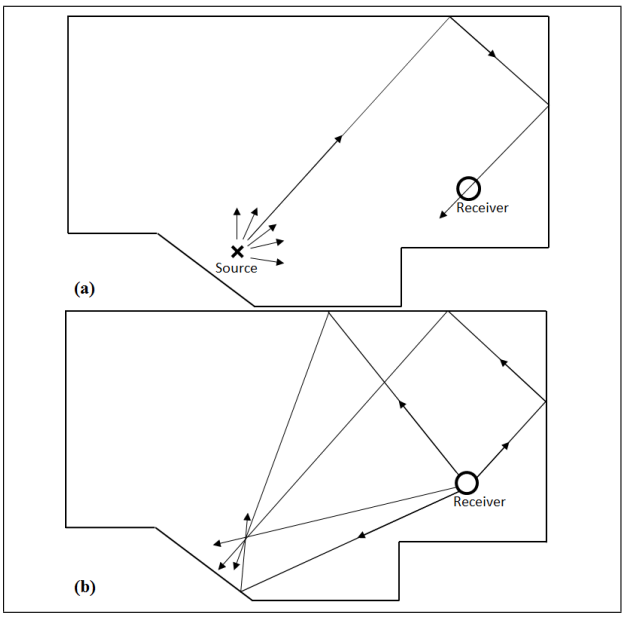


Figure 3. (a) Tracing a ray path between source and receiver. (b) Tracing ray paths from the receiver to estimate source positions.

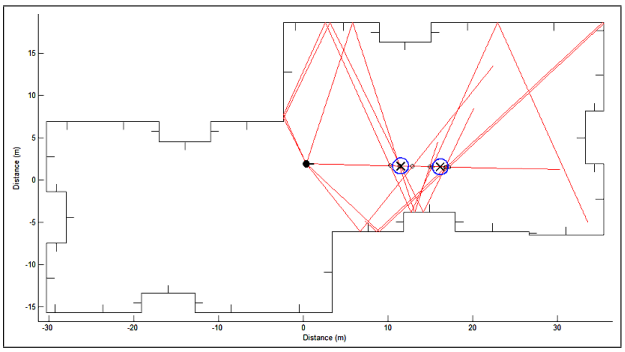


Figure 4. Screen-shot showing the result of tracing a set of rays extracted from a recording at position S1R1. The circular black marker indicates the receiver position and orientation. The red lines are the plotted ray paths. The black crosses mark the centre of clusters representing candidate source positions, and the blue circle around each black cross represents the mean data point to centroid distance in each cluster.

### 3.2.1. Source Position Calculation

Once all of the rays within a data set have been fired, the system calculates all of the possible source coordinates. This is achieved by checking for intersection between each section of the direct sound path and each section of every other plotted path. This is based on the assumption that the source will lie on the direct sound path at some position. Once all of these intersection points have been calculated, the system takes them and applies a $k$-means clustering algorithm [8] in order to identify the source position. Figure 4 shows a screen-shot of the localisation algorithm where a set of rays have been traced and, via $k$-means clustering (here $k = 2$), two candidate source positions have been identified.

# 4. EVALUATION

The localisation performance of the algorithm can be evaluated by measuring the distance between the system's best guess of a source's position and its true location. This section first considers how parameters might be optimised to give best performance. The performance of the algorithm is then evaluated, including results for different sound sources, different numbers of recordings used, and the effect of breaking the line-of-sight between source and receiver. This section concludes with results from applying the algorithm to a highly reverberant indoor space.

## 4.1. PARAMETERS

Parameters that will alter the early information extracted from the IRs include SIRR analysis window and hop size, and the frequency range over which sound intensity vectors are calculated. Once a set of early reflection information has been extracted, the reflection order to which ray paths are plotted will also have an effect on performance.

### 4.1.1. SIRR ANALYSIS WINDOWING

The window size defines the frequency resolution of the SIRR analysis, and hop size defines the time resolution. In order to avoid time smearing, the window size must be smaller or equal to the hop size used. Experimentation with various values found that a hop and window size of 512 samples gave the best compromise between time and frequency resolution. As such all the results included later in this section are based on SIRR analysis performed with hop and window sizes taking this value.

### 4.1.2. FREQUENCY RANGE

Table I shows the effect of changing SIRR analysis frequency range on localisation performance. These results indicate that the key directional information lies between 3 and 5 kHz. The degradation to performance when only considering frequencies below this range could be due to noise interference, as well as errors associated with modelling low frequencies with geometric modelling techniques. The poor performance resulting from the inclusion of frequencies above 5 kHz is most likely due to noise. All of the results presented later are the result of SIRR analysis performed over a frequency range of 500 Hz - 5 kHz.

### 4.1.3. REFLECTION ORDER

The maximum reflection order plotted for each path can have a significant effect on localisation performance. Table II shows results for localisation using recordings made at positions S1R1 and S3R1 given different maximum reflection order.

These results indicate an increasingly successful localisation performance with increasing reflection order up to third order. Beyond this order there ceases to be any real benefit, and in some cases results in a degradation to performance. All subsequently presented results were generated by plotting ray paths up to third order reflections.

Table I. Distance (in metres) between the ray tracer's best guess and the true position of source S1 for each recording method given different SIRR analysis frequency ranges (the minimum frequency in each case was 500 Hz).

| Max. Frequency (kHz) | Sine Sweep | Starter Pistol |
|---|---|---|
| 2 | 1.02 | 2.92 |
| 3 | 0.86 | 1.24 |
| 4 | 0.57 | 1.04 |
| 5 | 0.21 | 0.83 |
| 8 | 0.55 | 0.86 |
| 10 | 0.76 | 1.20 |

Table II. Distance (in metres) between the ray tracer's best guess and the true position of source S1 (from position S1R1), and S3 (from position S3R1) for differing maximum reflection orders plotted.

| S1R1 | | |
|---|---|---|
| Max. Reflection Order | Sine Sweep | Starter Pistol |
| 1 | 0.22 | 2.25 |
| 2 | 0.22 | 1.12 |
| 3 | 0.21 | 0.83 |
| 4 | 0.21 | 1.28 |
| **S3R1** | | |
| Max. Reflection Order | Sine Sweep | Starter Pistol |
| 1 | 8.24 | 15.08 |
| 2 | 3.63 | 8.31 |
| 3 | 2.79 | 5.49 |
| 4 | 2.79 | 5.49 |

Table III. Distance (in metres) between the ray tracer's best guess and the true position for each recording method at each recording position.

| Receiver Position | Sine Sweep | Starter Pistol |
|---|---|---|
| S1R1 | 0.21 | 0.83 |
| S1R2 | 0.95 | 1.14 |
| S1R3 | 0.93 | 0.37 |
| S2R1 | 1.93 | 3.58 |
| S2R2 | 1.19 | 4.09 |
| S2R3 | 2.75 | 1.89 |
| S3R1 | 2.79 | 5.49 |

## 4.2. RESULTS

Table III shows results for 'best guess' measurement indicating the distance between the estimated and actual source positions. The source and receiver positions are modelled as single points in space. In reality the transducers used are larger and non-ideal, with, for instance, the dimensions of the Genelec S30D being $495 \times 320 \times 290 mm$. For all results the 'true' source position is considered as a point source at the centre of its horizontal profile.

Table IV. Re-expression of the absolute error measurements included in Table III as a percentage error with respect to the true direct path distance between source and receiver in each case.

| Receiver Position | Sine Sweep | Starter Pistol |
|---|---|---|
| S1R1 | 1.30 | 0.50 |
| S1R2 | 3.20 | 3.80 |
| S1R3 | 2.13 | 0.85 |
| S2R1 | 8.87 | 16.44 |
| S2R2 | 3.26 | 11.21 |
| S2R3 | 5.27 | 3.63 |
| S3R1 | 5.58 | 10.98 |

#### 4.2.1. Sound Source

The results in Table III show the sine sweep method to give better localisation results than the starter pistol, most likely due to the high signal-to-noise ratio of the exponential sine sweep method and the consistency of its SNR across a large range of frequencies [9].

#### 4.2.2. Distance

Table III show the absolute distance between the algorithm's best guess of the source position and the true position in each case, exhibiting a decrease in localisation accuracy with increasing source-receiver distance as a result of increased transmission loss due to scattering and absorption. Table IV re-expresses these results as a percentage of the direct sound path in each case. These results still show a degradation to performance with greater source-receiver distance, but not to the extent suggested by Table III.

#### 4.2.3. Non Line-Of-Sight

The consideration of specular reflections alone in the algorithm leads to the expectation that localisation performance will suffer for scenarios where diffracted sound waves may contain key spatial information; for example where there is no direct line-of-sight between source and receiver.

The results in Tables III-IV align with this expectation, indicating generally less successful localisation for the non line-of-sight source/receiver configuration when using a starter pistol, especially when using starter pistol recordings. This is most likely due to the low SNR associated with the starter pistol relative to the exponential sine sweep method.

#### 4.2.4. Number Of Recordings

The recording of multiple IRs at several receiver position for sources S1 and S2 allows for evaluation of the algorithm performance, given different numbers of early reflection data sets used as the input. In the case of source S1 this includes recordings at S1R1, S1R2, and S1R3 (S2R1-S2R3 for source S2).

Table V. Distance (in metres) between the ray tracer's best guess, and the true position, resulting from the use of between 1 and 3 input data sets, for source positions S1 and S2.

| Source S1 | | |
|---|---|---|
| Data Sets Used | Sine Sweep | Starter Pistol |
| 1 | 0.21 | 0.37 |
| 2 | 0.09 | 0.23 |
| 3 | 0.20 | 1.57 |
| **Source S2** | | |
| Data Sets Used | Sine Sweep | Starter Pistol |
| 1 | 1.19 | 1.89 |
| 2 | 1.39 | 1.81 |
| 3 | 1.36 | 2.13 |

Figure 5. External view of St Margaret's Church (NCEM) [10]

The results in Table V show the algorithm's best guess of source position from firing one ray set, the best guess from the three possible combinations of two ray sets (e.g. the best from using S1R1 and S1R2, S1R1 and S1R3, or S1R2 and S1R3), and given the use of all three available data sets in each case.

The results in Table V do not indicate any clear correlation between an increase in source localisation accuracy and the number of data sets used. This could be in part due to minor variation in receiver orientation between recordings.

#### 4.2.5. Localisation in a Bounded Space

In order to further test the capabilities of the localisation algorithm, its performance has been evaluated in a highly reverberant and enclosed environment, totally different from that which the algorithm was designed for. The chosen location was the National Centre for Early Music (NCEM) at St. Margaret's Church, York, UK (Figure 5).

When St. Margaret's was redeveloped to become the NCEM, acoustic absorption panels and drapes in the ceiling were added to allow the acoustic behaviour to be changed for different settings (concerts, lectures, etc.). Recent work at the University of York generating auralisations of the church resulted in the generation of a computer model of the space, and a set of IRs recorded at several position in the church with the panels configured in each of three variations [11]. A full set of these results

Table VI. Results for localisation using the NCEM data. Each measurement represents the distance between system's best guess of the source position, and the true source position (in metres).

| Receiver Position | 1st Con. | 2nd Con. | 3rd Con. |
|---|---|---|---|
| R1 | 8.56 | 5.37 | 5.95 |
| R18 | 6.69 | 1.78 | 9.85 |
| R23 | 1.35 | 1.83 | 1.53 |

Table VII. Results for localisation using the NCEM data. Here the measurements are expressed as a percentage of the distance between source and receiver in each case.

| Receiver Position | 1st Con. | 2nd Con. | 3rd Con. |
|---|---|---|---|
| R1 | 77.30 | 48.80 | 54.09 |
| R18 | 69.89 | 18.57 | 102.60 |
| R23 | 22.50 | 30.45 | 25.50 |

and details of source positions are available on OpenAIR at [12].

Tables VI-VII show localisation performance expressed as absolute and relative distance respectively for three recording positions (R1, R18, and R23 - details of these positions can be found at [12]) at the NCEM, with a recording made at each position for each of three acoustic panel configurations.

The first configuration, designed for opera performances, makes use of the ceiling drapes and 75% of the panel absorbers. The second configuration also makes use of the drapes and 100% of the absorbers, and is suitable for lectures. The third configuration only uses the ceiling drapes, with all of the panels in the closed positions, and is for use in music recitals.

The results in Tables VI-VII expose the limitations faced by the developed localisation algorithm when applied to a bounded environment. The lack of identifiable early reflections means that the extracted information is erroneous, degrading the localisation performance significantly.

## 5. Conclusion

This paper covered the design and implementation of a source localisation algorithm suitable for sparse open environments exhibiting distinct early reflections in their IRs. Early reflection information extracted from B-format IRs is plotted on a 2D model of the area and localisation is performed using the intersection points of the plotted ray paths.

Results for IRs recorded in a semi-enclosed courtyard indicate the suitability of the algorithm to such an environment, with the inevitable introduction of errors by assumption of 2D propagation not creating any apparent issues. Localisation accuracy was shown to decrease with greater source-receiver distance, and the breaking of line-of-sight between source and receiver.

An evaluation of the algorithm's performance in a bounded, highly reverberant, environment exposed the algorithm's limitations. Without clear early reflections distinct from the reverberant section of the IR, the localisation performance was not successful.

### Acknowledgement

This project has been supported by an EPSRC internship in collaboration with Roke Manor Research.

## References

[1] F. Stevens, D.T. Murphy: Spatial Impulse Response Measurement In An Urban Environment. Proc. AES 55th Int. Conference, 2014.

[2] F. Stevens: OpenAIR - Alcuin College, University of York. `http://www.openairlib.net/auralizationdb/content/alcuin-college-university-york`, 2014.

[3] D.T. Murphy, S. Shelley, A. Chadwick: B-format Acoustic Impulse Response Measurement And Analysis In The Forest At Koli National Park, Finland. Proc. DAFx-13, 2013.

[4] J. Merimaa, V. Pulkki: Spatial Impulse Response Rendering I: Analysis And Synthesis. J. Audio Eng. Soc, 52(3) 115-1127, 2005.

[5] T. Korhonen, S. Tervo, T. Lokki: Estimation Of Reflections From Impulse Responses. Proc. International Symposium on Room Acoustics, 2010.

[6] X. Serra: A System For Sound Analysis/Transformation/Synthesis Based On A Deterministic Plus Stochastic Decomposition. PhD Thesis, 1989.

[7] S. Tervo. Direction Estimation Based On Sound Intensity Vectors. Proc. 17th European Signal Processing Conference, 2009.

[8] T. Moon, W. Stirling: Mathematical Methods And Algorithms For Signal Processing. Prentice Hall, 2000.

[9] A. Farina. Advancements In Impulse Response Measurements By Sine Sweeps. Proc. 122nd AES Convention, 2007.

[10] NCEM: The National Centre for Early Music `http://www.ncem.co.uk`, 2015.

[11] A. Foteinou, D.T. Murphy: Multi-Positional Acoustics Measurements For Auralization of St Margaret's Church, York, UK. Proc. Forum Acusticum 2014.

[12] A. Foteinou: OpenAIR - St. Margaret's Church. `http://www.openairlib.net/auralizationdb/content/st-margarets-church-national-centre-early-music`, 2011.