Determination and Display of Spatial and Temporal Room Impulse Responses

Daniel Protheroe, Bernard Guillemin

Department of Electrical and Computer Engineering, The University of Auckland, Auckland, New Zealand

A paper previously presented at ISSA 2010, 29-31 August 2010, Auckland

Abstract

The acoustical characteristics of a room are traditionally determined using directionless impulse response measurements, the directionless nature of which imposes significant limitations on the kinds of information that can be extracted. Obtaining such information in 3D would permit, for instance, a more detailed analysis of sound reflections to be undertaken. This paper overviews a technique for capturing a room's impulse response in 3D using a practical mi-crophone array (A-format) and then transforming this data into a Cartesian coordinate system (B-format) so that it can be visualised as an image or video. This 3D impulse response visualisation carries information about sound reflections in terms of sound intensity, spatial orientation and time. A software package has been developed in MATLAB to implement this technique, preliminary results from which will be presented. Enhancements of this 3D visualisation, together with a thorough understanding of the accuracy and robustness of the technique, are the focus of a current study and will be briefly discussed.

Introduction

The acoustical characteristics of a room are typically measured by recording the room's impulse response at a number of locations. An excitation signal that represents an impulse function, often in the form of a swept sinusoid, is generated at some specified location in a room (e.g., the stage in a concert hall). This signal is then recorded at one of a set of repre-sentative listening positions, and mathematically converted into a set of impulse responses. The impulse response at a specific location describes how the room reflects and absorbs sound waves as observed at that location. It is traditionally recorded using an omnidirectional pressure microphone, with the result that all spatial information is combined into a single audio signal and a directional analysis of the room reflections observed at that location is not possible. Obtaining the impulse response in 3D would permit an acoustician to identify the direction of specific reflections, and relate this information to the physical features of the room under study.

In this current work, a special microphone array is used to record a set of impulse responses. A software application is then used to extract information from these recordings and display it as an image.

An example of such an image, generated from a 3D impulse response, is shown in Figure 1. This image is made up of a number of lines, where each line corresponds to energy re-corded at the measurement location as a function of magnitude, orientation and time of arrival. The length of a line represents magnitude, the orientation of a line corresponds to the angle of incidence, and the colour of a line represents time of arrival. The scale is represented in Decibels, and the axes convey spatial direction.

As part of the current study alternative ways of representing the acoustical data are being explored, together with the ro-bustness and accuracy of the analysis routines used to extract it.

The next section explains some concepts which are funda-mental to the techniques discussed in this paper. Subsequently a system for determining and visualising 3D room impulse responses is described. This is in terms of a brief overview followed by an in

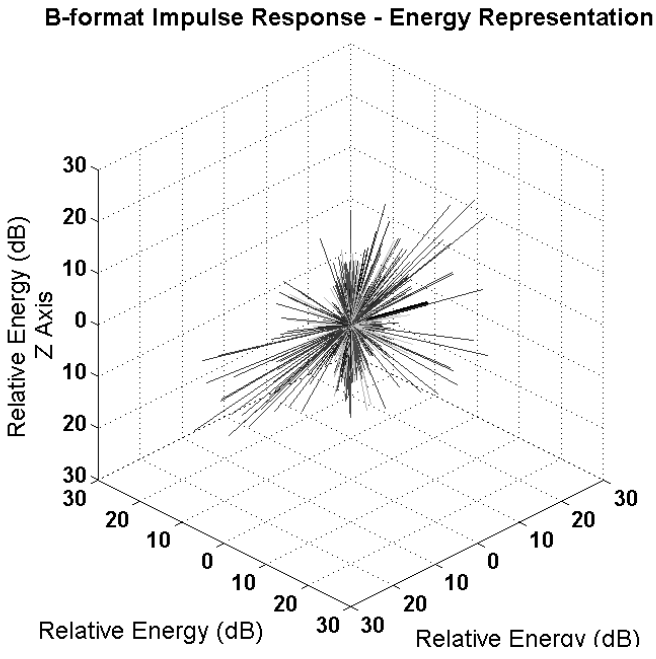


Figure 1: Example of a 3D impulse response image. Source: [1]

depth discussion of each part of the system. Currently the system is being experimentally validated in a controlled environment in order to assess its accuracy and robustness. An example of one such validation experiment is presented, followed by a discussion of planned future experimental work. Enhancements to the visualisation of the room impulse response are also discussed, followed by some concluding statements.

Background Fundamentals

The concepts of sound field, sound intensity, active and reactive sound fields, and B-format and A-format representations of a sound field are discussed in this section, as they underpin the techniques to be described later in this paper.

Sound Field Properties

A vibrating object, such as a loudspeaker, will cause mechanical disturbances in the surrounding medium. Similar to the effect of dropping a rock into a pond, the disturbances will propagate as sound waves outwards from the vibrating object, called the source, into a physical region in which sound waves exist, called the sound field.

A sound source causes the disturbance from steady state of several physical quantities of a medium. In the context of this paper, the medium of the sound field is air and the steady state refers to the time-invariant, ambient air conditions. Disturbance of this medium refers to the perturbation of the scalar quantities of pressure and density, as well as the oscillatory motion of the air particles which can be described by vector quantities such as displacement, velocity or acceleration. For a complete objective description of a sound field at a particular position, it is necessary to know both the scalar and vector aspects of these quantities. In this context the most useful and commonly used quantities are pressure and particle velocity. Of these, the fluctuating pressure (called the sound wave pressure), over the frequency range between 20 Hz and 20 kHz, is the most important for a subjective analysis, as this is the quantity that the human hearing system responds to.

Sound Intensity

From the standpoint of propagation of energy, a sound source radiates mechanical energy and the resulting propagating waves are a transmission of that energy through the sound field. Sound intensity (W/m2) is a quantity that describes the rate of energy flow through a unit area in the sound field. It is a vector quantity because it carries information about direction and is therefore one parameter used in the directional analysis of a sound field.

Instantaneous sound intensity, I(t), as distinct from sound intensity, I, at a position in the sound field, is defined as the product of the instantaneous sound wave pressure p(t) and the instantaneous particle velocity vector u(t) [2]:

$$I = p(t)u(t) \qquad (1)$$

The particle velocity vector, and thus the intensity vector, at a position in a sound field are inherently 3D quantities.

Sound intensity is then a time averaged rate of energy flow and is defined as:

$$\overline{I(t)=p(t)u(t)} \qquad (2)$$

where the bar indicates a suitable length time average [2]. At a position in a sound field where energy is flowing back and forth, the net flow or net intensity will be zero.

Active and Reactive Sound Fields

A purely active sound field is one in which there is a net energy flow, an example being a plane wave propagating in a free field (i.e., a sound field with no reflections). Conversely, in a completely reactive sound field, all the energy is travelling back and forth and the net energy flow is zero. For most sound fields, there will be both active and reactive components [3].

The active sound intensity, measured at a position in a sound field, is the time averaged sound intensity mentioned previ-ously. The reactive sound intensity gives an indication of the back and forth flow of energy through that position [4].

B-Format and A-Format Representations of a Sound Field

One way of representing a sound field at a position in terms of magnitude and direction is in the so called B-format. This is described in terms of sound wave pressure and 3D particle velocity. Sound wave pressure is a scalar quantity. Particle velocity is measured in a Cartesian coordinate system (broken down into its orthogonal X, Y and Z components), as shown in Figure 2(a). The positive directions in respect to this X, Y and Z coordinate system are forward, left and upwards, respectively. In terms of B-format, the pressure signal is referred to as the W channel, and the X, Y and Z components of the particle velocity are referred to as the X, Y and Z channels, respectively [5] [6].

One approach to determining the particle velocity experimen-tally would be to use an appropriately configured set of ve-locity microphones aligned in the X, Y and Z directions. Each of these microphones would need to have a directional response that resembled a figure-of-eight, as shown in Figure 3. The W channel, corresponding to sound wave pressure, would be determined by an omnidirectional microphone placed at that same location, as also shown in Figure 3.

In practice, however, it is very difficult to position and align microphones in the way that has been described. An alternative approach to solving this problem, developed by Michael Gerzon in 1975 [8], is based upon the A-format representation of a sound field. The coordinate system associated with A-format can be visualised in terms of the four vertices of a cube surrounding the measurement position. These four vertices are referred to as FLU, FRD, BLD and BRU, corresponding to front-left-up, front-right-down, etc., as shown in Figure 2(b). Unlike B-format, obtaining experimentally the A-format representation of a sound field at a location is relatively straightforward through the use of an A-format microphone array. Four directional cardioid microphones, which are partially sensitive to sound wave pressure, and partially sensitive to particle velocity, are aligned in the four A-format directions, facing outwards from the measurement position. One such microphone array is the Core Sound TetraMic [9], shown in Figure 4.

The signals produced by the A-format microphone array may be combined and filtered in such a way as to obtain the

equivalent B-format representation, as will be discussed in the next section.

System Overview

This section briefly outlines the system used in this study to transform from a measured impulse response (IR), determined using an A-format microphone array, into a 3D visualisation of this response. The functional blocks of this system are shown in Figure 5.

Firstly, the impulse response is recorded in terms of the A-format representation of a sound field by means of an A-format microphone array. The resulting set of signals is then converted into B-format. Once in this form, the sound intensity impulse response signals are calculated, these carrying information about both direct and reflected sound energy at the measurement location. This information is characterised in terms of magnitude, direction and time, and can then be visualised in terms of vectors on a 3D Cartesian coordinate system.

The following sections discuss each part of this system in greater detail. (Note: the details concerning the generation of impulse signals and the recording of impulse responses are not discussed in this paper).

A-Format To B-Format Conversion

In this section, a method is described to convert from A-format to B-format signal recordings. The theoretical framework underpinning this conversion process was initially formulated by Michael Gerzon [8]. Improvements to this process intended to enhance the accuracy and robustness of the technique have subsequently been

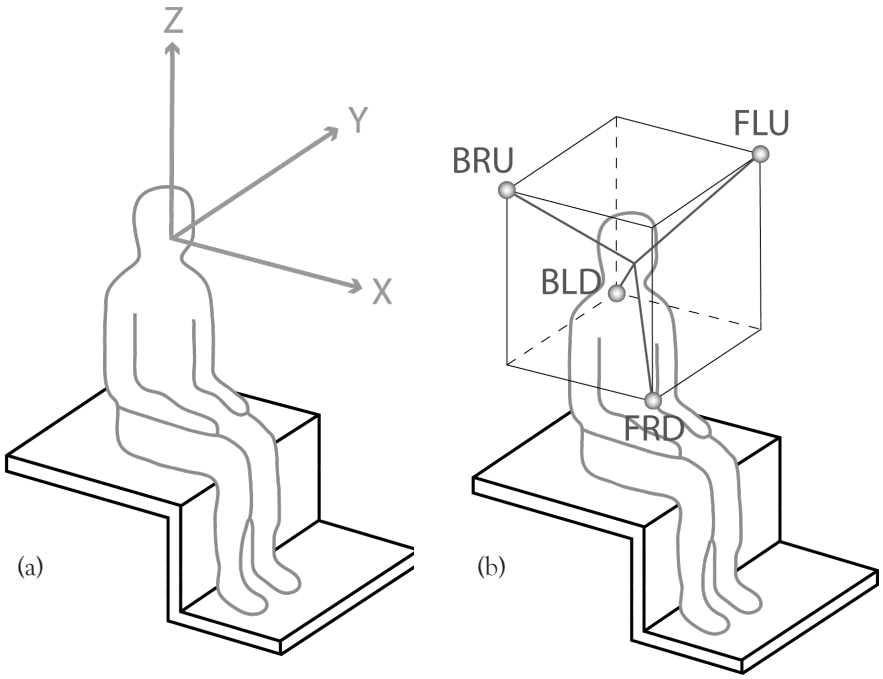


Figure 2: (a) B-format Cartesian coordinate system (b) A-format coordinate system. Adapted from [5].

presented by Angelo Farina [5].

A block diagram of Gerzon's theoretical framework is pre-sented in Figure 6.

The fundamental part of this conversion process is a matrix operation on the four A-format signals, as defined in Equation 3:

$$W' = FLU + FRD + BLD + BRU$$

$$X' = FLU + FRD - BLD - BRU$$

$$Y' = FLU - FRD + BLD - BRU$$

$$Z' = FLU - FRD - BLD + BRU \qquad (3)$$

The four signals at the output of this matrix operation are essentially in B-format but still require some post-filtering. Thus they are denoted with a prime signal. Post-filtering is required as a result of the non-coincident effects associated with the placement of the microphones which causes the signals, after the matrix operation, to have directivity patterns that deviate from the ideal [6] [10]. (Note: The ideal patterns were shown in Figure 3.) Gerzon derived theoretical post-filters to account for these issues, but it has subsequently been shown that these filters do not provide a realistic correction [5] [6]. A major reason for this is due to the fact that Gerzon's theoretical filters assumed a perfectly ideal microphone array. In reality, however, each microphone will exhibit non-ideal frequency response and directional characteristics.

Farina proposed an enhancement of Gerzon's approach, in the form of a set of pre-filters, as shown in Figure 7, in order to overcome the limitations just described.

These pre-filters equalise the frequency

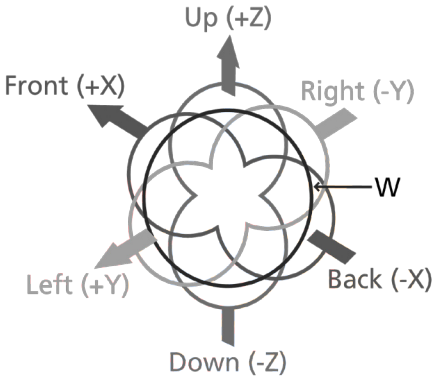


Figure 3: B-format directivity patterns, centred at the measurement point. Source: Adapted from [7].

response of each A-format microphone capsule, thereby ensuring that each capsule responds identically to the others. The filters are determined using an empirical method based on impulse response measurements as described by Farina [5].

Sound Intensity Analysis

Once the impulse response

Figure 4: Core Sound TetraMic A-format microphone array Source: [9].

measurements have been converted from A-format to B-format, a sound intensity analysis can be undertaken, providing information on direct and reflected sound energy in terms of magnitude, direction and time of arrival.

The B-format signals represent the sound wave pressure and 3D particle velocity at a position in a sound field. Therefore, computing the 3D sound intensity is straightforward, as defined in Equation 2. The result is three signals: the sound intensities in the X, Y and Z directions. For example, the sound intensity calculation for the X direction is shown in Equation 4. The intensity and particle velocity in the X direction are $I_X(t)$, and $u_X(t)$, respectively. The bar indicates a suitable length time average, as will be discussed further below.

$$I_X(t) = \overline{p(t)u_X(t)} \quad (4)$$

In undertaking this sound intensity analysis, in order to get information about sound reflections from objects in the room, it is necessary to distinguish between energy flow related to direct sound, early reflections and subsequent reverberation, as shown in Figure 8.

The rationale underpinning the classification of recorded sound intensity into these three classes will now be discussed.

In a typical room, the sound field will be composed of time varying active and reactive components, and knowledge of these properties will determine what directional information can be obtained from the sound intensity impulse responses. Firstly, it is assumed that for room measurements the microphone is located in the far field of the source (i.e., not immediately close to the source) and thus reactive near-field effects do not exist.

When only the direct sound is propagating in the sound field, as far as any particular location is concerned, there will be a net energy flow in one direction with respect to the source and thus the intensity at that location will be entirely active.

As soon as sound waves are reflected by objects in the room, multiple sound propagation paths of varying directions will result within the room. However, at any position in the far field, the early reflections at that position should be spaced relatively far apart in time, and hence the individual reflections should be clearly identifiable within the impulse response as shown in Figure 8. For the duration of the direct sound and each early reflection, there is energy flow in generally one direction and thus the intensity is mostly active. As far as the time average associated with determining $I_X(t)$, $I_Y(t)$ and $I_Z(t)$ is concerned (see for example Equation 4), this is chosen so that it does not overlap multiple reflections.

If two early reflections of different directions occur at a similar time (i.e., within this sound intensity time average) in some location, the measured intensity will be partially active and partially reactive, depending on the direction of the sound waves. The resulting analysis will then be an error.

As time progresses, the reflection paths will become more complicated, and the observed reflections at a position will merge together in time (see reverberation section in Figure 8). Eventually there will be reflections in many different and opposing directions at the same time, or within the same sound intensity time average. The active intensity will become zero, as the sound field is now mostly reactive (called a diffuse field).

The process of firstly distinguishing between early reflections as compared to reverberations and then subsequently deciding upon the temporal positions over which individual time averages will be performed is one that is largely guided by a careful examination of the overall impulse response. In many cases, the reverberation section will be largely

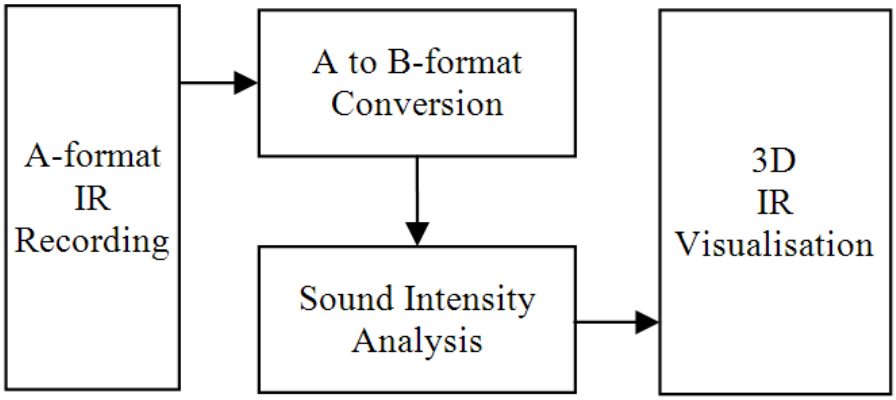


Figure 5: System Overview.

characterised by magnitudes that are smaller on average than those in the early reflection period. As far as the early reflection period is concerned, this will need to get segmented into subintervals over which individual time averages will be performed. Identifying these subintervals is largely done by identifying individual peaks in the early reflection period, the rationale being that each one of these peaks will correspond to an individual reflection. It should be clear from this explanation that the process of determining where individual time averages will be performed is not one that can be automated easily.

The X, Y and Z values for each time average will produce the magnitude and direction information for a single vector on the 3D impulse response image, as shown, for example, in Figure 1. The time information for a particular vector on the 3D impulse response image is governed by where in time the time average has been taken.

Visualisation Of 3D Impulse Response

The idea of visualising the 3D impulse response of a sound field is not new [11] [12].

The visualisation method presented in this paper generates an image in a similar style to the 3D impulse response visualisations by Alban Bassuet [12]. Firstly, the time-averaged sound intensity vectors for the direct sound and individual reflections are determined, as discussed in the previous section. The length of each vector (i.e., the magnitude of each reflection) is represented on a Decibel scale, termed the Relative Sound Intensity Level. The vectors are then plotted as lines on 3D axes, in a similar style to Figure 1. The time information associated with each vector is represented in terms of colour. With the current system the resulting 3D image plot can be rotated and viewed from any angle; an invaluable feature for room analysis. Alternative ways of looking at this information are currently being investigated as discussed in a later section.

Experimental Validation

This section outlines the experimental

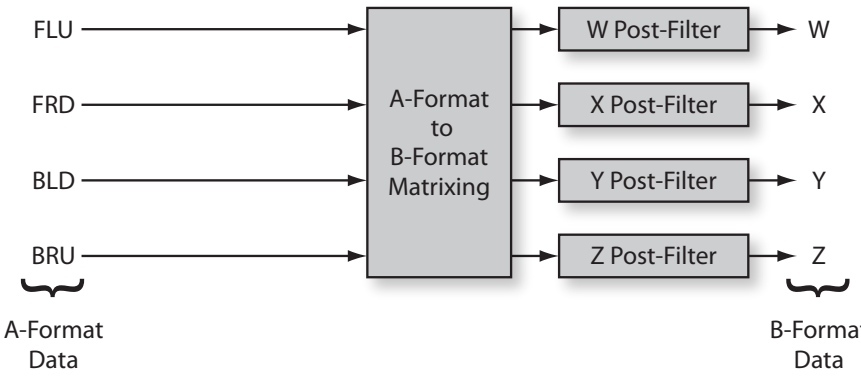


Figure 6: Gerzon's method for converting from A-format to B-format Adapted from [8].

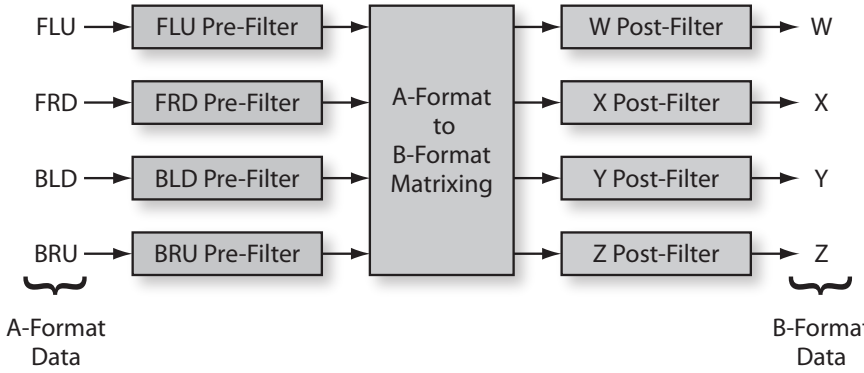


Figure 7: Farina's method for converting from A-format to B-format Adapted from [5].

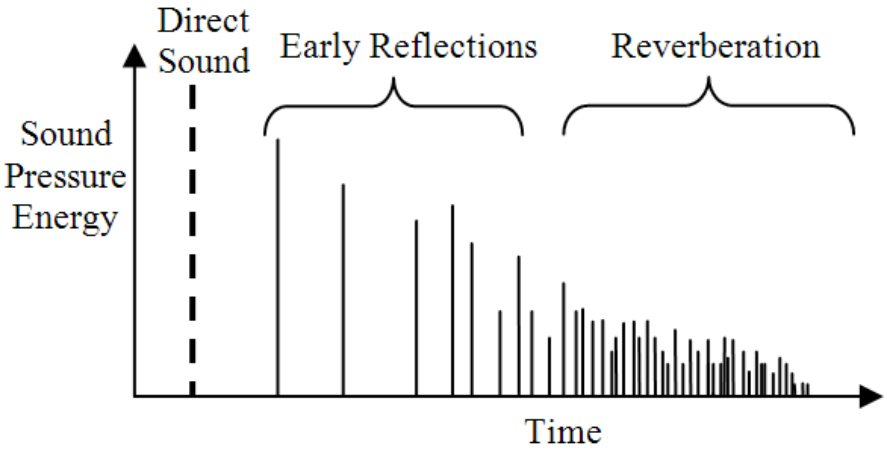


Figure 8: Illustration of simple omnidirectional energy impulse response at a position within a room.

environment which is currently being used to investigate the accuracy and robustness of the analysis process described in this paper. For practitioners, it is very important that these aspects are thoroughly understood before the technique is employed in a more realistic, and thus complicated, environment. The facilities used in this study to investigate these aspects of accuracy and robustness are described next. The results from one such experiment will be presented, followed by a discussion of the kinds of experimental validation that is planned.

Facilities and Equipment

As part of this study, experiments are being undertaken in an anechoic room with internal dimensions of "5m"×"5m"×"5m" . This facility exists within the Acoustics Research Centre of The University of Auckland. Figure 9 shows a photograph of an experimental setup inside this anechoic room. This figure shows an A-format microphone

array (Core Sound TetraMic) together with a sound source (Tapco S5 loudspeaker).

An Apple Macbook Pro computer, connected to a Metric Halo ULN8 multichannel audio interface is used to play back the excitation signals, and also record the four channels from the A-format microphone array. This equipment is located in a control room, adjacent to the anechoic room. The software FuzzMeasure Pro is used to generate a swept sinusoid excitation signal to be played back through the loudspeaker. This software also converts the recorded microphone signals into impulse response signals. Included with the TetraMic micro-phone array is software called VVMic which implements the A-format to B-format conversion process. Having obtained a set of B-format impulse responses, these are then plotted in MATLAB in order to identify the early reflection period as distinct from reverberation, as well as the temporal locations of the individual time averages within the early reflection period. MATLAB is then again used to compute the time-averaged sound intensity vectors for the individual reflections, as well as to visualise this data interactively on Cartesian axes in the form shown in Figure 1.

Example of a Typical Experiment

Figure 10 shows a typical experiment conducted to assess the accuracy and robustness of the techniques described in this paper.

With this experiment, there was only one reflector placed some distance away from both the loudspeaker and A-format microphone array. The TetraMic A-format microphone array provides an indication of the positive X direction in terms of the B-format coordinate system. This positive X direction was then made to point directly to the loudspeaker, as shown in Figure 10. The positive Y direction follows automatically from this, as also indicated in Figure 10, with the positive Z direction pointing out of the page. As far as the vertical placement of the reflector is concerned, it is located in the negative Z direction with respect to the microphone position. The blue arrows in Figure 10 are the expected paths of sound propagation from the loudspeaker, and the red arrow represents the expected reflection that

Figure 9: Anechoic room with a TetraMic A-format micro-phone array and Tapco loudspeakerresponse at a position within a room.

should be recorded by the microphone array. The angle ϑ (Figure 10) was physically determined using a laser measuring device to be approximately 112°.

The impulse response associated with the W channel for this experiment is shown in Figure 11. In this figure the direct sound and reflected sound are clearly obvious as there are no other reflection sources or noise present in the room.

The X-Y plane associated with the resulting 3D plot for this experiment is shown in Figure 12. The origin of this plot represents the location of the microphone. The vector coloured blue, in the near horizontal direction, is the direct sound intensity vector, and the reflected sound intensity is shown as the vector in red.

Examination of this plot shows the direct sound vector is offset by an angle of 2° in the X-Y plane. The angle between the direct and reflected sound is approximately 126°. This is an error of 14° from the angle ϑ shown in Figure 10.

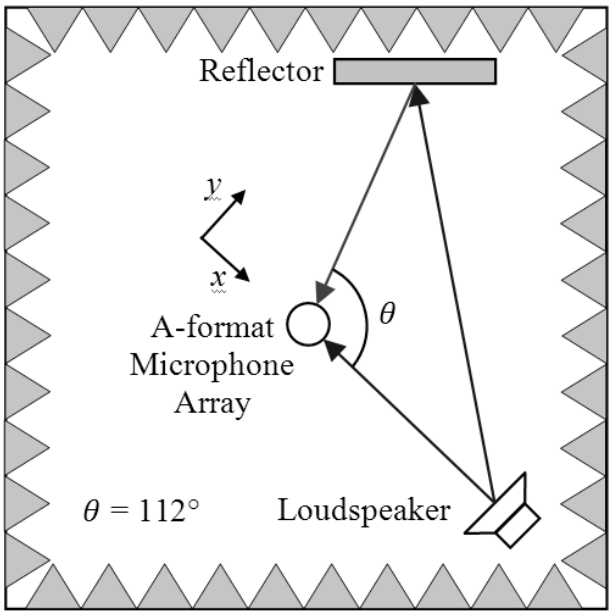


Figure 10: Experimental setup in the anechoic room with a loudspeaker, A-format microphone and reflector.

The angular error associated with the direct sound intensity vector seems very acceptable given the inevitable errors associated with aligning the microphone with the speaker. The angular error in respect to ϑ could possibly be explained in two ways. Firstly, accurate placement of the reflector in respect to the XYZ coordinate system is difficult in practice. Secondly, the nature of the reflection produced by the reflector used in this experiment is not fully understood and will be one of the aspects examined in further investigation. Nonetheless, the results associated with this initial experiment are very promising.

The sound intensity vectors can be visualised in 3D, as illustrated in Figure 13. This represents the same data as Figure 12, but now including the Z dimension. The reflection vector (red), with respect to the origin, is in fact pointing away from the page and downwards, although this is not obvious from the figure. This is consistent with the location of the reflector, which in this case, was slightly lower than the height of the microphone.

This section outlines the methodology of future testing that is planned in order to quantify the accuracy and robustness of the system.

The experiment just described will be extended to include a single reflector placed at a variety of locations in respect to both angle ϑ (Figure 10) and separation distance from both the microphone and loudspeaker. This will then be extended to include a range of experiments involving multiple reflectors.

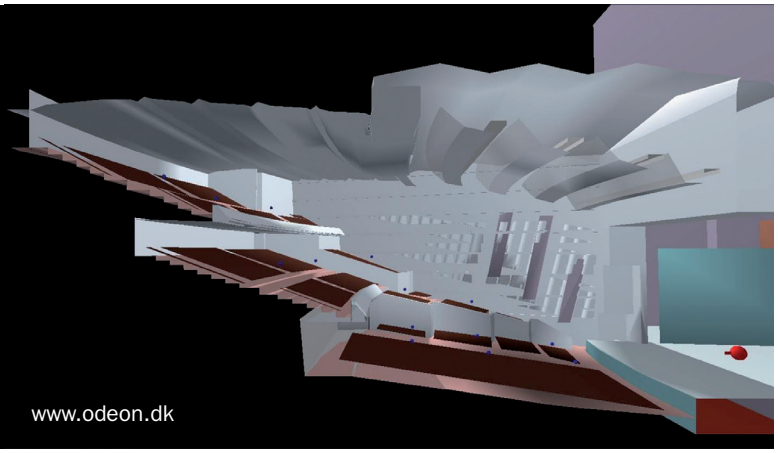
There are also issues associated with the compensation of the frequency response (i.e., both magnitude and phase response) of the loudspeaker as well as the recording equipment used in these experiments. Though it is expected that the resulting effects on accuracy will be small, it will nonetheless be useful to quantify these in some manner.

Following these experiments, the plan is to test this system in a more complicated environment such as a concert hall. In such a situation, both identifying the early reflections as distinct from reverberations, as well as deciding on the temporal locations for individual time averages associated with this analysis is likely to be non-trivial. The issue of trying to undertake this process in a more automated manner also needs investigation.

Enhancements to 3D Visualisation

In the aforementioned experiment, the sound field was very simple, consisting of only one reflection. In reality the sound fields will involve many reflections, and the resulting visualisation may look confusing, especially in terms of the temporal locations of the reflections. The ability to view only a specific window of time may also help to clarify understanding of the sound field with respect to time. Furthermore, rendering the visualisation as a video will help to show how the sound field changes over time.

The current method of visualising sound reflections as vectors on 3D axes is of some practical value, allowing acousticians to relate reflections to the corresponding physical features of the room. However this will require a manual process of matching the



intensity vectors to the room's surfaces. In order to simplify this task, the reflections could be viewed from the point of view of the measurement location. This will require a photograph to be taken from the measurement position in an appropriate direction.

The individual reflections will then be visually indicated in the estimated locations on the photograph, immediately relating the features of the room to the respective reflections. This idea was published in a previous paper, specifically relating to sound intensity measurements in enclosed spaces [13].

The visualisation may also be enhanced in terms of a perceptual analysis. Before calculating the sound intensity vectors, it may be appropriate to filter the signals into certain frequency bands which are of perceptual relevance.

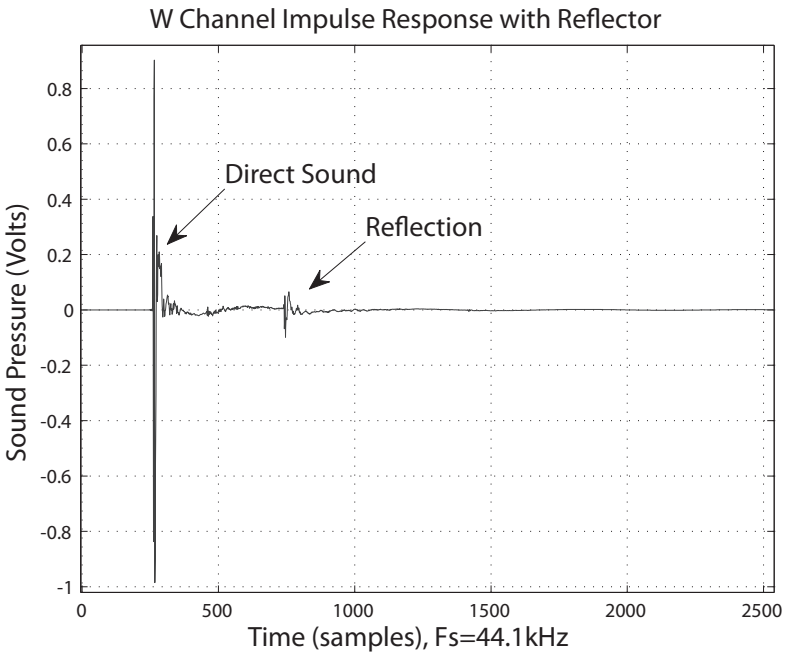


Figure 11: W channel (pressure) impulse response.

Conclusions

This paper has overviewed a technique for determining and displaying 3D impulse responses based upon measurements from an A-format microphone array. The intention with this technique is to be able to identify the magnitude, direction and time of arrival of reflections arising from various reflective surfaces of a room. The purpose of the current study is to ascertain the accuracy and robustness of this technique, as well as to identify the various factors which impact upon these parameters. An experimental setup has been described for undertaking this investigation based upon experiments conducted within an anechoic room. One such experiment has been described, the results of which are certainly promising in terms of the potential usefulness of this system.

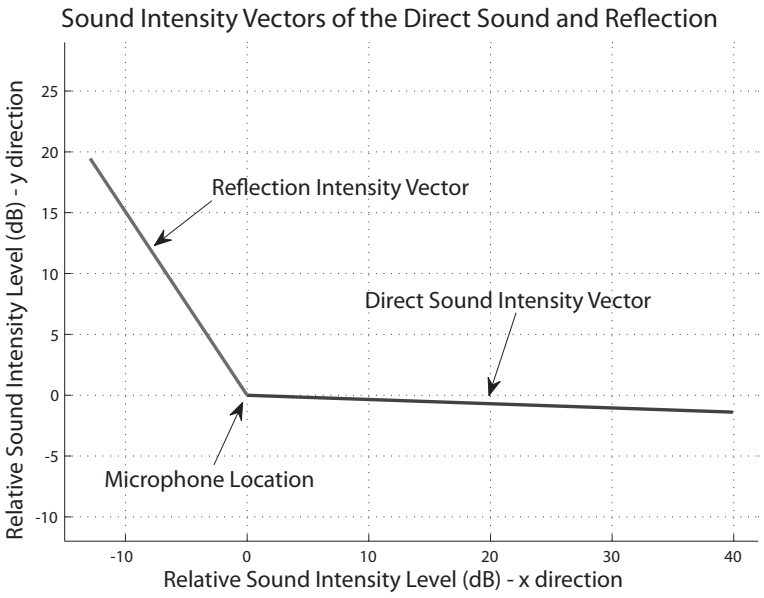


Figure 12: Sound intensity vectors in the X-Y plane of the direct sound and reflection.

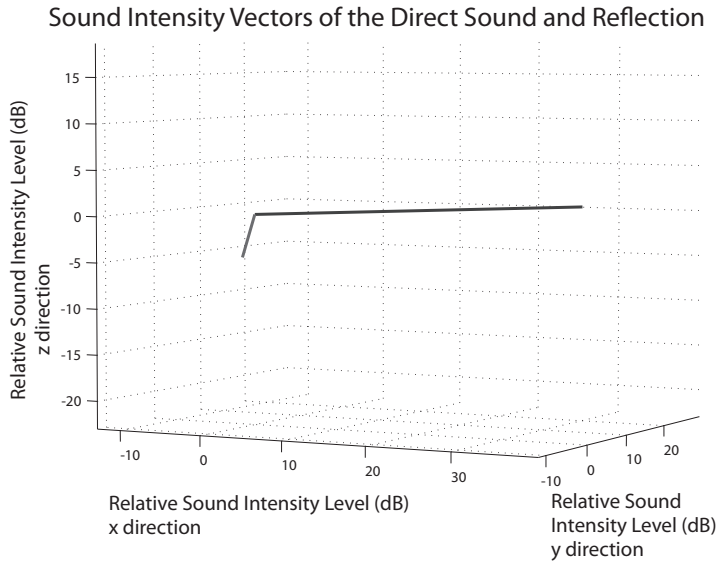


Figure 13: Sound intensity vectors visualised in 3D.

Acknowledgments

The authors would like to thank Keith Ballagh and the staff of Marshall Day Acoustics, Dr. George Dodd and Gian Schmid from the Acoustics Research Centre at the University of Auckland, and Prof. Angelo Farina of the University of Parma in Italy for their invaluable assistance and guidance.

References

1 D. Protheroe, "Development of a DSP application for the 3D representation of acoustic impulse responses," Department of Electrical and Computer Engineering, University of Auckland, Final Year Undergraduate Project Report 2009.

2 S. Gade, "Sound Intensity (Part 1 Theory)," Technical Review, Brüel & Kjær, no. 3, 1982.

3 Brüel & Kjær. (1993) Primer: Sound Intensity. [Online]. http://www.bksv.com/Library/Primers.aspx

4 F. J. Fahy, Sound Intensity, 2nd ed. London, UK: E & FN Spon, 1995.

5 A. Farina. (2006, October) A-Format to B-Format Conversion. [Online]. http://pcfarina.eng.unipr.it/Public/B-format/A2B-conversion/A2B.htm

6 F. Adriaensen, "A Tetrahedral Microphone Processor for Ambisonic Recording," in Proc. of the 5th Int. Linux Audio Conference (LAC07), Berlin, Germany, 2007, pp. 64-69.

7 Paul White. (2004, June) Recording A Live Choral Performance. [Online]. http://www.soundonsound.com/sos/jun04/articles/liveconcert.htm

8 M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," in 50th Audio Engineering Society Convention, London, 1975, Mathematical Institute, University of Oxford, England.

9 Core Sound. (2009) TetraMic. [Online]. http://www.core-sound.com/TetraMic/1.php

10 C. Faller and M. Kolundija, "Design and Limitations of Non-Coincidence Correction Filters for Soundfield Microphones," in 126th Audio Engineering Society Convention, Germany, 2009.

11 A. Abdou and R. W. Guy, "Spatial information of sound fields for room-acoustics evaluation and diagnosis," Journal of the Acoustical Society of America, vol. 100, no. 5, pp. 3215-3226, November 1996.

12 A. Bassuet, "Acoustics of a selection of famous 18th century opera houses: Versailles, Markgräfliches, Drottningholm, Schwetzingen," in Acoustics 08 Paris, Paris, 2008, pp. 1645-1650.

13 Y. Fukushima, H. Suzuki, and A. Omoto, "Visualization of reflected sound in enclosed space by sound intensity measurement," Acoustical Science and Technology, vol. 27, no. 3, pp. 187-189, 2006. ¶.