CHARACTERISATION OF SMALL ROOM ACOUSTICS FOR AUDIO PRODUCTION.

J. Romero School of Computing & Engineering, University of Huddersfield, U.K.
B. Fazenda School of Computing & Engineering, University of Huddersfield, U.K.
H. Atmoko School of Computing & Engineering, University of Huddersfield, U.K.

1 INTRODUCTION.

Nowadays there is an explosion of home music production activities such as recording, mixing and mastering, which usually takes place in small spaces, which are not ideal for such tasks. Their reduced volume and the typical construction structure of homes, which usually have the common factor of strong walls facilitates the formation of standing waves and non diffuse sound decays which decrease the immersion sense and the realism of the program. The interference of the sound field with the relatively big size of the furniture and equipment in relation with the dimensions of the room affects also the quality of sound.

Coloration of sound perceived as a change of timbre, rhythm sensation and signal[1] pitch is the main problem encountered in music reproduced in these spaces. It happens varying in a complicated way, as it depends on temporal and spatial variables. In order to minimize these problems, acoustic treatment needs to be applied to the room. The traditional ways of measuring rooms to assess this treatment had not taken in account the directionality of the sound energy and their effect in terms of perception of stereo image without blur estimation caused by harmful reflections, neither had taken in account what are the characteristics of neutral rooms.

This document aims to investigate the extraction and analysis of temporal and spatial distribution early sound decay. A method based on B-format signals is adapted to small rooms. It can map the spatial and temporal distribution of sound energy and diffuseness in three dimensions. Once the data is collected it is possible to extract the information in time and frequency domain and use it to infer issues of perception based on psychoacoustic models. The ultimate objective is to find out useful descriptors to characterize the acoustic quality in critical listening rooms.

2 SPATIAL MEASUREMENT TECHNIQUES FOR SMALL ROOMS

According to theory, there are three basic types of measurements: a) Separate microphone arrays, using discrete microphone arrays where the phase relationship of signals is not preserved correctly b) Coincident microphone arrays which minimize the distance between the capsules in order to converge to a point in the space. An example of this is the *Soundfield* microphone, which was used for the measurements. c) Intensity probe configurations (p-p probes) use a pair of measurement grade pressure calibrated capsules placed at fixed distances. The signals are processed to obtain the acoustic intensity vector along the axis where the probe is aligned.

Further advances had been reported with a new approach based in transducers sensitive to particle velocity ($\vec{u}$) known as the *Microflown* probe[2]. This solution drastically minimizes the distance between the transducers and is therefore currently the best coincident array, showing the most reliable polar pattern across frequency and the less intrusive probe up to date.

The work of Merimaa[3] and Peltonen[4] settled the standards of extraction of directional impulse responses and diffuseness analysis with the use of perceptual issues applied to the frequency

content on the signals. Later Pulkki[5] and Ahonen[6] expanded its uses to teleconferencing. Enroth[7] applied these techniques for auralization of simulated acoustic impulse responses to be reproduced accurately in multiple channels. Essert[8] developed a method based on the *Soundfield* microphone to measure the Lateral Energy Fraction (LEF), the Instantaneous Lateral Fraction and the Directional Fraction to describe the evolution of the sound field for hall design. Ohta[9] analysed reflections using a time windowing approach extracting B-Format signals from a custom microphone made of six cardioids capsules.

Fazenda[10] used B-Format signals to extract directional distribution of reflections across time. However poor localization was found at angles near 45º when the instantaneous particle velocity vector was calculated in the time domain. Therefore the motivation of using active Intensity and Short Fourier Transform ($STFT$) approach was implemented in the present work. It should also be noted that Faller[11] and Farina[12] reported a limitation of the *Soundfield* microphone for accurate estimation of reflections due to the lack of exact first order polar patterns across the entire audible range reported as asymmetries and distortion of the original directivity patterns. Above 12 kHz phase errors are present, the polar directivity of the figure of eight and the omnidirectional patterns are distorted.

The $STFT$ analysis is used in the present work to obtain better estimation of reflection directionality than when it is computed in time domain. It also helps to visualize the data in two dimensions at the same time. The downside of this approach is that once the window time resolution is set, the frequency resolution depends on the number of samples contained in the time window. Longer time windows have enhanced frequency resolution but poorer time resolution. Using the highest sampling rate helps to minimise the aliasing of frequencies which happen slightly at a low and more noticeable at a high frequencies. This constraint means that a fine resolution in both domains cannot be achieved at the same time and that is a problem for the visualization of the same data because only one variable can be focused at a given time. It is only with experience gained using these techniques that one can select the most suitable resolution. The more advanced transforms i.e. Winger-Ville transform can help to overcome this limitation at an expense of more computational power, which limits its use on real time measurement devices.

3 DESCRIPTION OF THE MEASUREMENT SYSTEM

The *Soundfield* SPS 422B microphone is used to extract 3-D impulse responses (www.soundfield.com). It has four cardioid capsules, which generate the A-Format signal (see figure 1). By applying a matrix algorithm within either a hardware unit, or software implementation, it is possible to obtain the B-Format signal. It consists of four simultaneous signals: an omnidirectional polar response signal (W), which represents the acoustic pressure (p) with unity gain. The other remainder three signals are orthogonal virtual figures of eight microphone polar responses centred at the origin. The Soundfield rack processor deliver the dipole Signals ($\vec{X}$, $\vec{Y}$, $\vec{Z}$) with a gain of $\sqrt{2}$ (+3 dB) referenced to the omnidirectional signal W. This is standardised in order to give the same level as the signal W in free field conditions. The reason is to maximize the signal to noise ratio for analogue recordings[13]. In this analysis the acoustic pressure is taken from the value of the W channel i.e. (p = W). For further analysis in *Matlab*, the dipole signals (X, Y, Z) which represent particle velocity vectors ($\vec{u}(n) = X(n)\hat{i} + Y(n)\hat{j} + Z(n)\hat{k}$) are downscaled by dividing them by root of two[14]:

$$
\begin{aligned}
W_{corrected}(n) &= W_{measured}(n) \\
\vec{X}_{corrected}(n) &= \frac{1}{\sqrt{2}}\left[\vec{X}_{measured}(n)\right] \\
\vec{Y}_{corrected}(n) &= \frac{1}{\sqrt{2}}\left[\vec{Y}_{measured}(n)\right] \\
\vec{Z}_{corrected}(n) &= \frac{1}{\sqrt{2}}\left[\vec{Z}_{measured}(n)\right]
\end{aligned}
\qquad (1)
$$

Where n is the discrete time variable after the analogue-digital converter and the arrow above the letter mean a vector quantity. This approach gives a better scale of diffuseness estimate where the lowest value (zero = purely specular reflection) yields a true white colour rather than using the initial approach suggested by Fazenda[10].

This type of measurements requires a calibrated *Sounfield* microphone and a reliable sample and phase accurate soundcard connected to a silent portable computer. Measurement software performs the extraction of the impulse response of the four B-Format signals. (Refer to figure 1). *Matlab* is used for further analysis discussed in Section 4. The excitation source was the existing studio monitors found in the rooms.

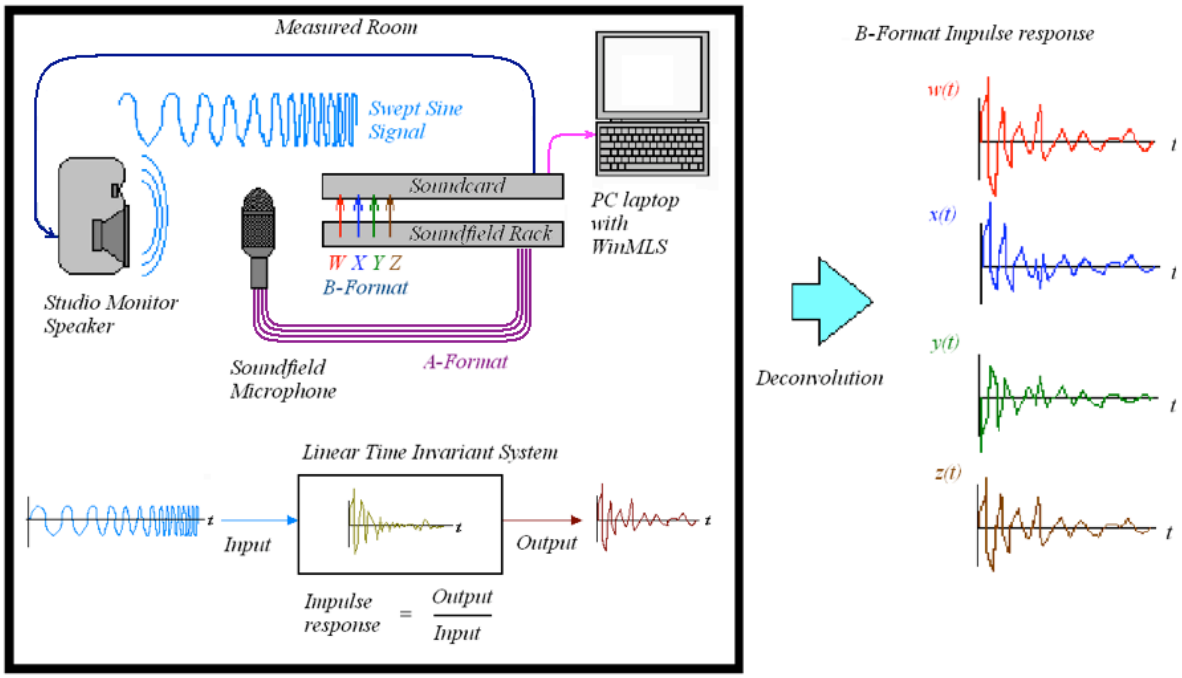

Figure 1: Measurement system diagram.

The impulse response of the room represents the interaction of the speaker and the room at this particular position. It is important to note that these impulse responses are not good to calculate any acoustic parameter due to the fact that the studio monitors are not omnidirectional at all frequencies. To obtain the impulse response, a deconvolution process is performed with the aid of the measurement software *WinMLS* (http://www.winmls.com). The impulse response from the measurement is generated in binary or `ACII` file formats. The excitation signal chosen was linear swept sine. The sweep length was selected according to the measured RT, to improve accuracy. The four impulse responses were obtained by performing simultaneous RT measurements by selecting the '4 channel mode', with the soundcard operating at its highest resolution which was a bit depth of 24 bits and a sampling rate of 96 kHz. The final step is to export data to *Matlab* for further analysis using `Loadimp.m`, a script provided by *WinMLS*. The B-Format signals are normalized with the maximum value of the four signals.

4 DATA ANALYSIS

This analysis involves two domains at the same instant. It can map the behaviour of a variable in time and frequency by using the short time Fourier transform ($STFT$). The first step is to divide the impulse response by a chosen number of time windows of length m and applying the fast Fourier transform to the product of each of the B-Format signals with a Hanning window (w_{Hann}). Each `fft`

has a length of N points. The time windows need to overlap in order to create a smooth transition between each window. Each time window is zero padded before and after it with half of the length of the time window and this is done to obtain the first sample multiplied with a window, which is at the maximum value to avoid spectral alias. The practical values of overlapping may be between 25 to 50 %. In order to create each slice of time, successive `fft` are applied to each component:

$$\begin{aligned} W(k) &= FFT\left\{W(n)\cdot w_{Hann}(n)\right\} \\ \vec{X}(k) &= FFT\left\{\vec{X}(n)\cdot w_{Hann}(n)\right\} \\ \vec{Y}(k) &= FFT\left\{\vec{Y}(n)\cdot w_{Hann}(n)\right\} \\ \vec{Z}(k) &= FFT\left\{\vec{Z}(n)\cdot w_{Hann}(n)\right\} \end{aligned} \tag{2}$$

Where k is the frequency index calculated by the `fft`. The three dimensional particle velocity vector is created by applying the short time Fourier transform ($SFFT$) to each orthogonal component and creating the following vector:

$$\vec{U}(n,k) = [X(n,k)\hat{i} + Y(n,k)\hat{j} + Z(n,k)\hat{k}] \tag{3}$$

Where n is the time index for each slice of time. The next step consists in calculating the intensity vector for each time window[13]:

$$\vec{I}_X(n,k) = \frac{\sqrt{2}}{\rho_0 c_0}\mathrm{Re}\left\{W^*(n,k)\vec{X}(n,k)\right\}\hat{i} \tag{4}$$

$$\vec{I}_Y(n,k) = \frac{\sqrt{2}}{\rho_0 c_0}\mathrm{Re}\left\{W^*(n,k)\vec{Y}(n,k)\right\}\hat{j} \tag{5}$$

$$\vec{I}_Z(n,k) = \frac{\sqrt{2}}{\rho_0 c_0}\mathrm{Re}\left\{W^*(n,k)\vec{Z}(n,k)\right\}\hat{k} \tag{6}$$

Where ρ_0 is the air density and c_0 is the speed of the sound, the orthogonal component unit vectors are $\hat{i}$, $\hat{j}$, $\hat{k}$ and * is the complex conjugated value of pressure signal W. The active intensity components are combined to create the three dimensional active intensity vector:

$$\vec{I}(n,k) = \left[I_X(n,k)\hat{i} + I_Y(n,k)\hat{j} + I_Z(n,k)\hat{k}\right] \tag{7}$$

According to Ahonen[6], the direction of arrival of the active intensity ($\bar{I}$) is opposite to the direction of the particle velocity vector $\vec{U}(n,k)$. The formulas read:

$$\bar{I}_{XY}(n,k) = \sqrt{\left(-\vec{I}_X(n,k)\right)^2 + \left(-\vec{I}_Y(n,k)\right)^2} \tag{8}$$

$$\theta_{\bar{I}_{XY}}(n,k) = \left(\frac{180^\circ}{\pi}\right)\tan^{-1}\left(\frac{-\vec{I}_Y(n,k)}{-\vec{I}_X(n,k)}\right) \tag{9}$$

$$\phi_{\bar{I}_{XZ}}(n,k) = \tan^{-1}\left(\frac{-\vec{I}_Z(n,k)}{\sqrt{\left(-\vec{I}_X(n,k)\right)^2 + \left(-\vec{I}_X(n,k)\right)^2}}\right) \tag{10}$$

The degree of diffuseness of a sound field is a scalar value, which measures the ratio between arriving directions of active intensity and energy density at a given time window. Pulkki[5] explains that diffuseness varies between values of 0, when the direction of the incoming reflections directions is clearly determined i.e. an isolated reflection and 1, when the direction of the incoming reflections is random (implying a diffuse field). In figures 4a and 4b darker grey spot values correspond to higher values and the white values to lower values of diffuseness.

Further development from applying the analysis to B-Format signals led to an instantaneous function of diffuseness (ψ) in terms of Short Term Fourier Transform windows ($STFT$) in a room by applying the following expression[7,13]:

$$\psi(n,k) = 1 - \left[\frac{\sqrt{2}\left\| \mathrm{Re}\left\{ W^{*}(n,k)\vec{U}(n,k) \right\} \right\|}{\left|W(n,k)\right|^{2} + \frac{1}{2}\left(\left|-\vec{I}_{X}(n,k)\right|^{2} + \left|-\vec{I}_{Y}(n,k)\right|^{2} + \left|-\vec{I}_{Z}(n,k)\right|^{2} \right)} \right] \qquad (11)$$

The instantaneous value of diffuseness is calculated by applying the concept of the 'Directional Diffusion Index' (Ψ) to the active intensity orthogonal component vectors calculated from B-format signal.

The *Matlab*'s '`norm`' operation is applied to the Intensity vector $\left\|\vec{I}(n,k)\right\| = \sqrt{\left(\vec{I}_{X}\right)^{2} + \left(\vec{I}_{Y}\right)^{2} + \left(\vec{I}_{Z}\right)^{2}}$ to the 'expected value' of intensity $\left\langle \vec{I}(n,k) \right\rangle$, which in practice is calculated with the recursive integration of the $STFT$ of the active intensity vector $\vec{I}(n,k)$.

In order to obtain the directions of the reflections, an $STFT$ analysis is made to the intensity vector in the horizontal plane. The time window duration is chosen according to the sampling frequency and the 'mean free path' of the room. Octave band, 1/3 octave band, or Equivalent Rectangular Band (ERB) filtering can be used to smooth the time windows. The best perceptual approximation is the ERB because it resembles the human hearing resolution, whereas 1/3 octave band approximates the critical band with less accuracy and octave band may be useful for engineering analysis. The formula used to calculate the bandwidth of the rectangular filters (ERB_N) is taken from Moore[14]:

$$ERB_{N} = 24.7\left(4.37 \times 10^{-3} f + 1\right) \qquad (12)$$

Where ERB_N and frequency (f) are in Hz. According to Venegas[15], the interaural spectral density per band between the left and the right signal is a special form of the $IACC$ and it can obtained after filtering the signal with a bank filter that resembles the function of basilar membrane. The time domain is divided in smaller angular regions. According to Blauert's localization blur[16] (see figure 2). By applying small time windows to the data it is possible to analyse the early part of the decay but not the late one. This is due to the fact that reflections become too dense and they are perceived as reverberation rather than isolated discrete reflections. Nevertheless, a statistical approach is still possible in the late decay zone using selective wider time windows to cover the entire reverberant tail.

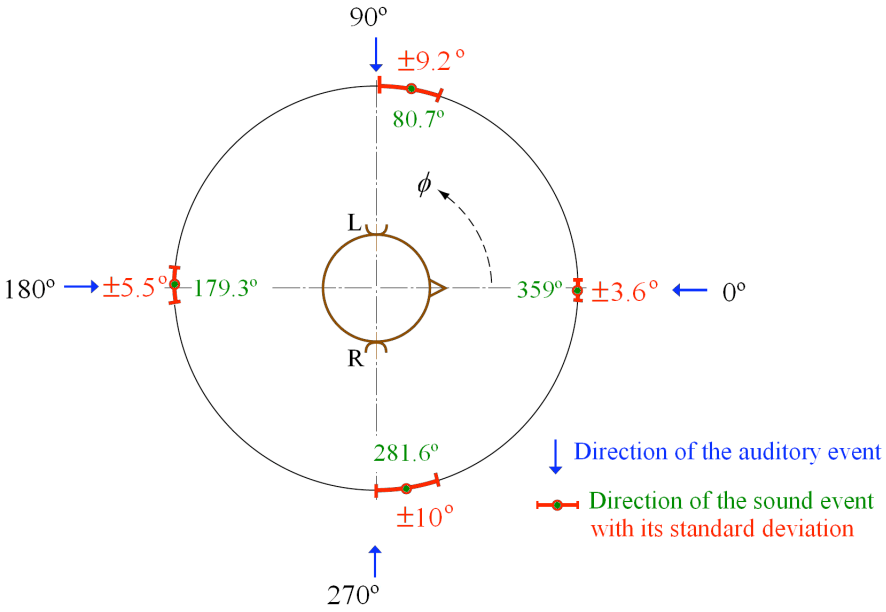

Figure 2: The map of localization blur in the horizontal plane. (after Blauert (1997)).

According to Buchholz[17], the perception of several room reflections is masked by the direct signal and in some cases by other loud reflections. By applying a threshold to discard the masked energy based on a psychoacoustic model is possible to take in account the relevant information. This model may take in account two variables, angle and time of arrival.

In this work it is necessary to assess the speed at which a room becomes diffuse. This has been attempted in two ways. Originally as suggested by Fazenda[10], the numbers of occurrences of reflections within an angular region of 20 degrees were counted. However with this approach the *SPL* level of reflections was not considered which has subsequently been shown to be of importance. Therefore a new method is proposed. It studies the evolution of the distribution of reflections across both time and frequency with respect to the localisation blur model. This second approach would seem more promising.

According to Geddes[18] diffraction and very early reflections (< 1 ms) are more important than spectral analysis because they are more prone to be heard because they are not masked and may be the cause of poor temporal resolution than frequency resolution. Diffraction effects are level dependent so they are perceived in a non-linear way. It is critical to control early reflections in the design of small rooms and the accepted design parameter is to avoid any early reflection within a time window of 15 ms referred to the direct sound and excluding the first order floor or ceiling reflections[20]. The purpose is to obtain an initial time delay gap *ITDG* figure, which can provide a true accurate stereo image not mixed with harmful early reflections.

Another factor to avoid is near field diffraction from the monitors, which contain sharp edges by flush mounting. Horns and waveguides in monitors at high levels produce noticeable diffraction due to sharp discontinuities in the path of the wave. Speakers need to have an omnidirectional directivity at all frequencies in order to have a natural recreation of the sound inside a room. Using omnidirectional sources usually creates the natural effects of reflections and the speakers need to approximate this attribute. Full bandwidth absorbers can be constructed inside the structure of the room. It may be possible to absorb low frequencies without taking the reverberance of a room by using membrane type absorbers made of MDF.

According to Geddes[19] the main difficulty of sound localization inside small rooms are the early lateral reflections off of the nearest walls. These reflections will be fused with the direct sound and

may bring confusion on the perception of location of virtual sources. They also cause sound coloration. The effect of the vertical reflections are not so significant on localization because our localization blur is poor in the median plane, but they can have a strong impact on coloration and timbre so a compromised treatment of ceilings with absorbers and diffusers is needed.

An early reflection arriving from opposite directions is good for creating a spatial impression due to the binaural cognitive hearing process. Nevertheless that is not the case for reflections, which arrives at the same ear. For this reason most reflections in the vertical plane are undesirable.

According to Blauert[16], image localization depends on the range of frequencies 1 kHz to 8 kHz and has higher sensitivity at the resonance of the ear canal around 3 kHz.

5 EXPERIMENTAL RESULTS MEASURED IN A CONTROL ROOM.

Only the horizontal plane is investigated at this time. The measurements were made in Blue Room 2 control Room, whose dimensions are 4.16 X 2.75 X 2.87 m and its volume is 32.1 m^3. Measurements were performed at two different source positions. The first was at the front of the microphone and the second at 120º from the origin, referenced at front with an approximated angle of 0º (which later was measured and was 2º, see figure 3). As previously mentioned, the sampling frequency resolution was set to 96 kHz which gives a temporal resolution of 0.01 ms. For simplicity, only one room measurement is presented here.

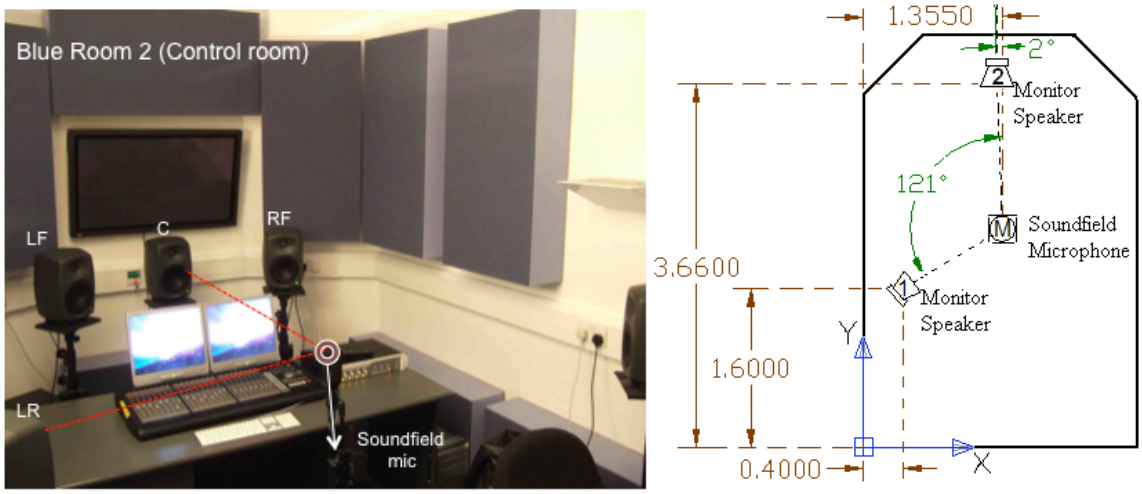

Figure 3: Photo of Blue room 2 control room (left) and schematic diagram of the measurement in meters (right).

This room has a fairly short reverberation time RT = 0.17 s @1 kHz, and a mean free path (MFP) $\bar{l}$ = 2.2 m giving a minimum resolution time window of $\Delta\bar{t}$ = 6 ms. However, smaller time windows of 0.1 and 1 ms were tested in this analysis and seem more appropriate to determine any isolated early reflection. For the $STFT$ a window length of N = 512 points and 50% overlap was used in order to produce a spectrogram (See figure 4a and 4b). The time scale begins at 0 ms, the region is adapted to observe the direct sound and its decay which was up to 30 ms. For the direct sound all the frequencies should have the same direction. For the first early reflections the direction of incoming should also follow this behaviour with some deviations, which can be caused by reactive impedances found in the boundaries of the room. The behaviour of reflective surfaces gives coherent frequency content on the reflections with good accuracy up to 5 kHz[13]. By analysing figure 4a and 4b in the upper graph the variation of direction of the early reflections is not traceable after the time labelled around 30 ms. This corresponds to the 10-15 ms limit where the reflections leave the discrete zone and increase its density across time. It is interesting to see that the more diffuse

the reflection, the less coherent the direction estimate is across the audible frequency range. In figures 4a and 4b it can be seen that the diffuseness values which appear as grey spots tend to vary randomly after the early reflections and therefore these are not useful descriptors of the stage of decay. One possible solution is to combine the diffuse estimate with the amplitude level of the reflections calculated with the Matlab's `quiver` graph. The direction estimations of reflections are calculated as the active intensity, which results from computing the mean direction of the reflections calculated after smoothing within each band of the *ERB* bank filter. From this analysis it is clear that new measures for taking in account the mixing time and the beginning of the reverberant tail are needed in order to perform an automatic analysis of each room measured. Possibly the use of higher order statistics and modern time series theory may be adapted to cover this needs.

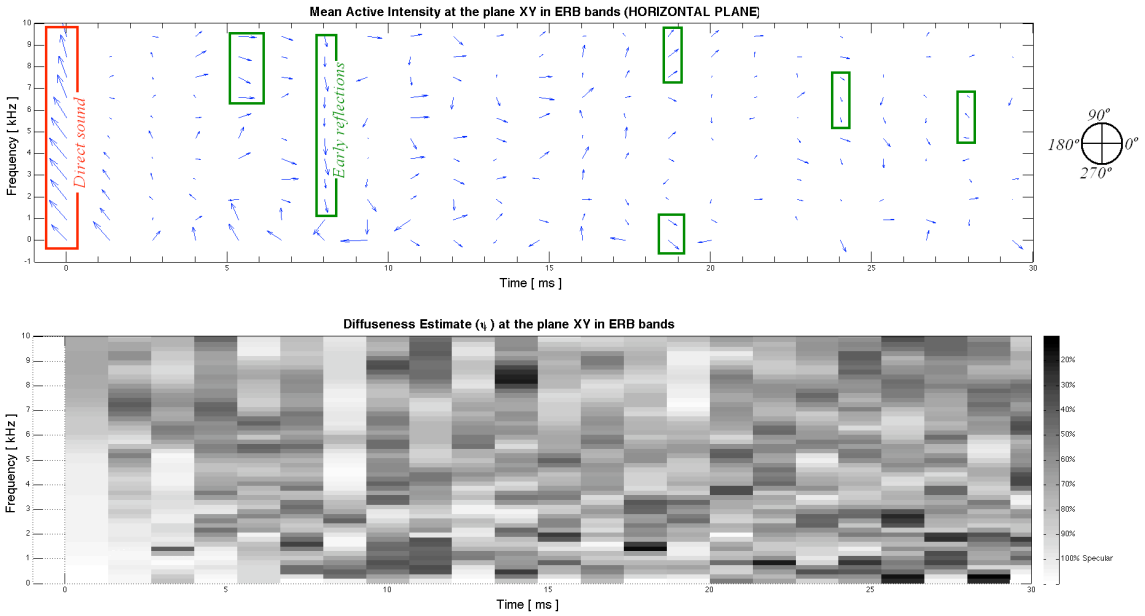

Figure 4a: Upper graph shows the direct, early reflections and late sound decay. Lower graph shows the diffuseness estimate (ψ) calculated with the *SFFT* method. The source was located in position 1. (It corresponds to the rear left side of the microphone from the sketch of figure 3). Note that the front direction is located at 0º in the present graph (which corresponds to an angle of -90º rotated from figure 3).

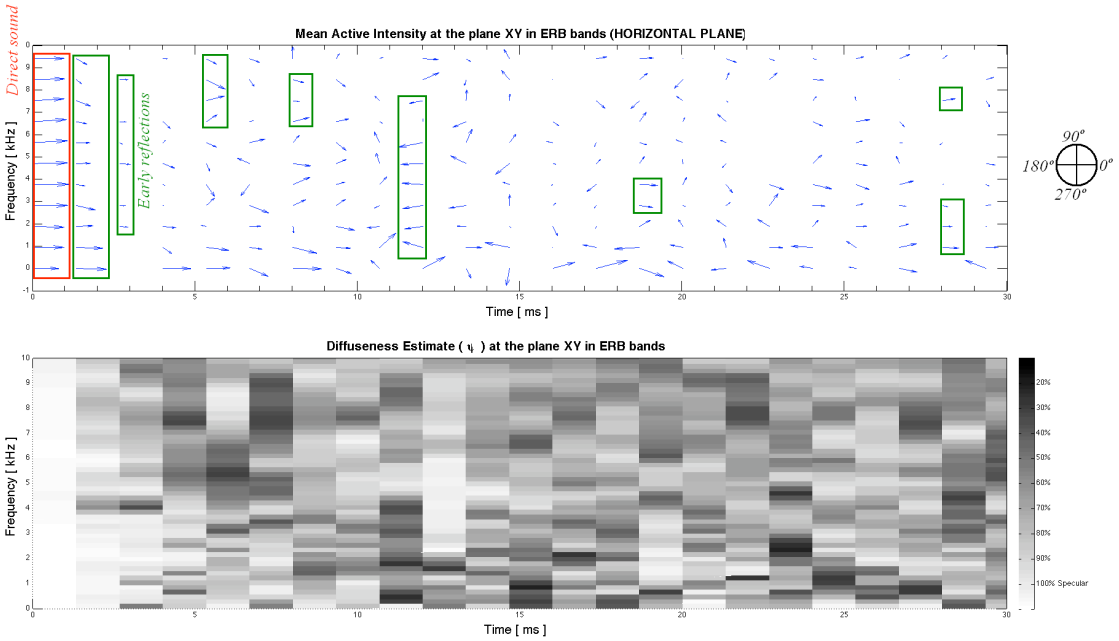

Figure 4b: Upper Graph shows the direct, early reflections and late sound decay. Lower graph shows the diffuseness estimate (ψ). The source used is located in position 2 (which corresponds to the front of the microphone. Note that the front direction is located at 0º in the present graph (which corresponds to an angle of -90º rotated from figure 3).

The direct sound is clearly detected in both figure 4a and 4b as the white colour on the diffuse estimate graph. It can be seen that the mean direction estimate for the direct sound on each frequency band is coherent in all the frequency bandwidth (red rectangles) and it tends to deviate slightly at frequencies above 8 kHz due to the limitation of the Soundfield correction filters. The model is capable of detecting very few full bandwidth reflections (green rectangles) in figure 4b and depending of the orientation of the source it can sense only one strong coherent reflections located around 13 ms. Due to this limitation in some respects, is preferable to use arbitrary chosen time windows to detect reflections as Ohta et al. had implemented before[9]. The reason of the appearance of some blur information which does not convey any reflection information may be the use of the overlapping window with 50 %, which may create some type of redundant information which is averaged twice because each window intersects common information and may create some ambiguous direction estimates. The choice of a stepper window than the long Hann window is compromised with the distortion, which may be found in its spectrum.

On the other hand the practical application of Fourier transform by the use of the matlab's `fft` creates a scaling problem which is partially solved by normalizing its instantaneous complex value dividing it by the length of the signal. However this approach sometimes need a careful review before combining the orthogonal components of vectors which may underestimate or overestimate spectral peaks. One problem found was that increasing the sampling resolution also creates the need to raise the length of the time window in order to maintain the same frequency resolution. That option is particularly not very useful for the measurement of short decay times. This is due to the fact that a lot of information is irrelevant because there is mainly background noise after and before the sound decay. The random behaviour of a signal may be filtered by subsequent averaging, but zero padding was found to be a complex procedure when the magnitudes need to be converted to dB and also it tends to deliver a wrong impression of the spectrum.

It is interesting to note that this type of directional analysis is not completely valid for measurements made only with one source-receiver position because its response varies considerably with different angles and different source positions. One possible reason is the fact that in small rooms only statistical measurements are meaningful[19].

6 CONCLUSIONS.

The estimation of direction of arrival of early reflections using the $STFT$ approach in time-frequency domain appears to be a good technique with high accuracy and repeatability in comparison with the time domain analysis, however both analyses are complementary. The inclusion of Equivalent Rectangular Band bank filter model helps bridge the gap between the objective measures and the subjective measures.

The physical limitation of the measuring system is the resolution of the *Soundfield* microphone, especially in the forwards and backward directions where the human localisation blur angle is more accurate. The future use of the *Microflown* probe may improve the measurement results, and better prepare the impulse response to create accurate auralizations.

7 REFERENCES

1. P. Rubak, L. G. Johansen, 'Coloration in natural and artificial room impulse responses' Presented at the 23rd AES International conference, Copenhagen, Denmark, May 23-25 (2003).
2. H. E. de Bree, 'An overview of Microflown Technologies', (2001). Available at: http://doc.utwente.nl/58819/1/Bree03overview.pdf [Accessed on August 31st 2009].
3. J. Merimaa, T. Lokki, T. Peltonen, M. Karjalainen 'Measurement, Analysis, and Visualization of Directional Room Responses', Presented at the AES 111th Convention, N.Y., U.S.A. September 21-24, (2001).
4. T. Peltonen, T. Lokki, B. Goutarbès, J. Merimaa, M. Karjalainen, 'A System for Multichannel and Binaural Room Response Measurements', Presented at the AES 110th Convention, Amsterdam, The Netherlands, May 12–15, (2001).
5. V. Pulkki, 'Spatial sound reproduction with directional audio coding', Journal of the Audio Engineering Society, Vol. 55 (6), pp. 503-516, (2007).
6. J. Ahonen, V. Pulkki, F. Kuech, M. Kallinger, R. Schultz-Amling, 'Directional analysis of sound field with linear microphone array and applications in sound reproduction' Presented at the AES 124th Convention, Amsterdam, The Netherlands, May 17-20, (2008).
7. S. Enroth, "Spatial Impulse Response Rendering of digital waveguide mesh room acoustic simulations", University of York, MSc Thesis, (2007).
8. R. Essert, 'Measurement of Spatial Impulse Responses with a Soundfield Microphone', Third joint meeting of the Acoustical Society of America and the Acoustical Society of Japan Honolulu, Hawaii, December 4-6 (1996). Available at: http://www.soundspacedesign.co.uk/1996a.pdf [Accessed on August 29th 2009].
9. T. Ohta, H. Yano, S. Yokoyama, H. Tachibana, 'Sound Source Localization by 3-D Sound Intensity Measurement using a 6-channel microphone system Part 2: Application in room acoustics', Internoise, 37th International Congress and Exposition on Noise Control Engineering, (2008).
10. B. Fazenda, J. Romero, '3-Dimensional Room Impulse response measurements in critical listening spaces', Proceedings of the Institute of Acoustics, Vol. 30 (6) pp. 323-239, (2008).
11. C. Faller, M. Kolundzija, 'Design and Limitations of Non-Coincidence Correction Filters for Soundfield Microphones', AES 126th Convention, Munich, Germany, May 7–10, (2009).
12 A. Farina, 'Advancements in impulse response measurements by sine sweeps', Presented at the AES 122nd Convention, Vienna, Austria, May 5–8, (2007).

13. J. Merimaa, “Analysis synthesis, and perception of spatial sound – Binaural localization modelling and multichannel loudspeaker reproduction” PhD Thesis, Helsinki University of Technology, Laboratory of Acoustics and Signal Processing ESPOO, page 21, (2006).

14. B.C. Moore, “An Introduction to the Psychology of Hearing”, Fifth Edition, Elsevier, U.K., (2004).

15. R. Venegas, M. Lara, R. Correa, S. Floody, ‘Spatial Sound Localization model using neural network’. Presented at the AES 120th Convention, Paris, France, May 20–23, (2006).

16. J. Blauert, “Spatial Hearing, The Psychophysics of Human Sound Localization”, Revised Edition, MIT Press, U.S.A. (1997).

17. J. M. Buchholz, J. Mourjopoulos, J. Blauert, ‘Room Masking: Understanding and Modelling the Masking of Room Reflections’, Presented at the AES 110th Conference, Amsterdam, the Netherlands, May 12-15, (2001).

18. E. R. Geddes, ‘Powerpoint presentation on Small Room Acoustics’, Presentation available at: http://www.gedlee.com/Papers.htm [accessed on August 29th 2009].

19. E. R. Geddes, “Premium Home Theater design and construction”, Chapter 4 Room Acoustics available at: http://www.gedlee.com/downloads/Chapter%204.pdf [Accessed on August 29th 2009].

20. M. Whitcroft, ‘Acoustic geometry sculptor - A computer program for optimizing room surfaces’, Proceedings of the Institute of Acoustics, Vol. 30 (6), pp. 41-50, (2008).