Audio Engineering Society

Convention Paper

Presented at the 142nd Convention
2017 May 20–23 Berlin, Germany

Quantitative investigation artificial room simulations reproduced by channel-based and object-based surround sound systems.

Bernard Camilleri[1], Jakob Bergner[1], and Christoph Sladeczek[1]

[1] *Fraunhofer Institute for Digital Media Technology IDMT, Ilmenau, Germany*

Correspondence should be addressed to Bernard Camilleri (camilleribernard@yahoo.co.uk)

ABSTRACT

The introduction of object-based audio reproduction comes along with new challenges for the sound engineer to record, design and synthesize reverberant sound fields due to the increased number of speakers and the placement of such. The aim of this paper is to show that several parameter settings from a digital reverberation unit produce contrasting reflectograms in a 5.0 channel-based setup and an object-based setup which can have effects on the perceived reverberant sound field. Conversely, established acoustical metrics derived from the measured room impulse responses (RIRs) in both multichannel reproduction setups do not highlight the differences noticed in the reflectograms. The potential consequences regarding individual system properties and the metrics themselves are discussed in this work.

1 Introduction

The investigation presented in this paper is part of the ongoing research on artificial acoustic room simulation in the field of spatial, object-based audio reproduction. The main objective of this work is to bring out the differences of artificial reverberant sound fields reproduced by an object-based audio reproduction system which is based on a scaled-down 2.5D Wave Field Synthesis (WFS) setup, compared to a standard channel-based 5.0 system. The comparison of the two systems is carried out by means of a Lexicon 960L digital reverberation unit while observing the temporal reflection patterns and energy characteristics of the resulting measured room impulse responses (RIRs) of both setups. The quantitative analysis is conducted by examining established room acoustical metrics including Early Decay Time (EDT), Reverberation Time (T30), bass ratio (BR), Clarity index (C80), Speech Transmission Index (STI), and Direct-to-reverberant ratio (DRR). This paper will focus on the EDT, T30 and D/R results.

Recent developments in the field of immersive spatial audio reproduction reached a market-ready status and thus became popular in various areas. Fields of applications include movie productions and cinema, virtual reality (VR), industrial virtual engineering and virtual acoustic environments (VAE) in general. Here different technologies and approaches are used with both academic and commercial use cases. The benefits concerning auditory immersion of technologies like wave field synthesis (WFS), higher order ambisonics (HOA), vector-based amplitude panning (VBAP) or binaural synthesis exceed the possibilities of traditional stereophonic and surround sound approaches like 5.1 or 7.1.

By introducing these new technological possibilities, the tools and workflows for content production in both artistic and realistic approaches must be adapted or even newly developed. This also includes the synthesis of artificial room acoustic behavior of sound. Established reverberation tools offer a myriad of possibilities to control the acoustic characteristics of a space (artificially synthesized or recorded), mainly adjusting the blend, size, tone, sustain and spread of a mix or individual instruments [1]. On basic two-channel and 3/2 stereophonic setups, the effects of reproduced reflections and the limitations of such setups are relatively known, especially when seeking to achieve spaciousness and envelopment in the listening area [2][3][4]. With the rise of above mentioned spatial audio reproduction technologies, the engineer has a greater palette of artistic possibilities to create a more realistic sound field, but at the same time can be faced with a rather complex task of designing or recording a reverberant environment to the requirements of the task at hand. As a result, engineers need to have a good understanding of how reverberation effects differ or conform in channel-based and object-based multichannel sound reproduction setups. This includes technical limits depending on the technology of choice as well as perceptive effects which will be discussed in this paper.

2 Background

2.1Object-based Audio

Today, most audio productions are still realized by using the channel-based audio approach. The sound engineer creates the spatial audio mix by modifying the gain of two or multiple loudspeakers for each audio signal. Known as panning, this technique allows the creation of so called phantom sources which can be perceived as coming from between adjacent loudspeakers [3]. When the panning is realized for the separate audio signals, the audio engineer is theoretically saving loudspeaker signals as a result of the spatial audio mix. One of the major drawbacks of this method is that this spatial audio mix is only valid for a single listener position called sweet spot as well as a dedicated loudspeaker setup (fixed loudspeaker position) is required for the best listening experience [3]. As both criterions cannot be fulfilled for most of the real life listening scenarios, spatial audio research focusses on optimizing this situation for a more immersive and realistic experience.

One of the approaches that came up in recent years is the concept of wave field synthesis [5]. Here the listening area is surrounded by (a theoretically infinite) loudspeaker array, in order to synthesize the sound field of virtual sources as accurately as possible. This is realized by calculating an individual driving signal for each speaker, so that the superposition of all loudspeaker signals form the wave field of an audio object. As the loudspeaker signals are calculated interactively, the result of an audio mix is no longer loudspeaker channels, but an audio object in a virtual space. An audio object is characterized by an input signal (e.g. a female voice, guitar, violin, etc…) and its corresponding meta data such as, position, gain or type of the audio object. Instead of playing back ready mixed loudspeaker channels, an audio renderer is transforming an audio object description into real loudspeaker signal. This approach is called object-based audio [6].

In this paper an object-based audio technology called SpatialSound Wave is used. The system is based on wave field synthesis theory but is additionally using modified algorithms in order to allow three-dimensional and less densely placed loudspeaker setups [6].

2.2 Room simulation in audio production

When listening to audio sources, room acoustics has a major influence on how a source is perceived [6], hence the popularity of room simulation tools in audio productions (apart from other creative uses). With the availability of object-based multichannel production systems one of the central questions is how to achieve this, mostly because we are dealing with a more realistic playback system which therefore requires even more realistic virtual room acoustic techniques. On the one hand there are new approaches based on sound field recording, simulation and synthesis techniques [**7**] [8]. On the other hand, only a few sound engineers have the possibility to use these complex algorithms, as the number of installations in audio production studios is still too little. Furthermore, there are a lot of well-established room simulation devices on the market which are well known and used by a lot of sound engineers. As these devices are mostly designed to create stereo or surround effects, one big question is how they can be used in object–based audio applications and what differences/similarities might they present to the sound engineer.

3 Methodology

The Methodology of this paper follows a strict comparative investigation of an object-based and a channel-based reproduction system. Measurement, processing, analysis and discussion is conducted with the same technical devices under constant conditions to reduce the degree of freedom to the respective reproduction system.The measurements of the resulting room impulse responses (RIR) were conducted in a listening room of 516m³ which fulfills the recommendation ITU-R BS.1116-1. The measurements were conducted by loudspeakers K&F CA106, microphone Gefell MK221 and software AFMG Easera.

3.1System under test

The reverberation unit used in this research to produce the synthesized reverberant room simulation was a Lexicon 960L. The 960L can take five discrete inputs and output five channels simultaneously (for Left (L), Right (R), Centre (C), Left Surround (LS) and Right Surround (RS) channels respectively), which makes it ideal for a production on a standard 5.1 channel-based surround sound setup. For measurements the center channel was chosen. The parameter variation includes temporal, spectral and artistic characteristics as shown in Table 1.

Test No.	Parameter	Value
1	'Dry' Reference test	100% dry
2	Initial Setting	*RT:* 1.477s, *size*: 25; 50% Wet
3	Increase in Size	40
4	Add Bass multiplier	1.9
5	Add reverb pre-delay	3ms
6	Add 'Contour'	+7
7	Change 'Contour'	-7
8	Add Early Delay level	29%
9	Add Early Delay master	56%
10	Add early delay (dLF>LF)	85.12ms
11	Add early delay (dRF>LF)	96.32ms
12	Add early delay (dRS>LS)	140ms
13	Increase Early Delay level	84%
14	Early reflection Roll-off	6.50kHz
15	Add reverb time Hi-cut	4.80kHz
16	Add 'Shape'	157
17	Add 'Spread'	168
18	Add 'Diffusion'	172
19	Add 'Spin'	62
20	Add 'Wander'	21.15

Table 1: Parameter variation of the reverberation unit Lexicon 960L

In this research the above mentioned object-based audio reproduction system SpatialSound Wave (SSW) was used. The setup under test is made up of 24 loudspeakers with an average spacing of 44cm set up in a circular array with a diameter of approximately 5m. The SSW software includes a mixer and a 'canvas' where virtual sources can be placed in relation to the physical loudspeaker setup.

As comparable reference a 5.0 surround setup according to ITU-775 recommendation was used on the same speaker layout.

The measurements of the RIRs were conducted at 7 different positions within the listening area for both setups under test. The loudspeaker and microphone setups for the channel-based and object-based system are depicted in Figure 1 and Figure 2 respectively.

3.2 Processing of measurements

To obtain the final results, some post-processing of the measured RIRs was necessary. The first step was to window the RIRs to remove excess noise and loudspeaker harmonic distortions which are 'pushed' towards the end of the RIR due to the deconvolution process of the logarithmic sweep [9]. For this, a symmetrical, rectangular time window was used so that the energy inside the window is kept intact. Additionally, the windowing ensures the correct calculation of the EDT and RT by removing the excess noise from the tail of the RIR, which would otherwise smear the Schroeder backward integration calculation [9] [10].

For an easier observation of the reflection patterns the RIRs were afterwards rectified and time-aligned to compensate the system delay of the SSW renderer.

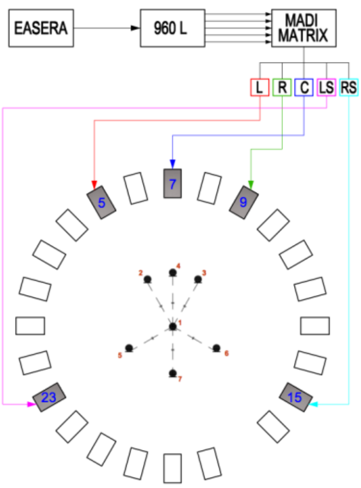

Figure 1: Loudspeaker array with chosen microphone positions for RIR measurements of the channel-based 5.1 system.

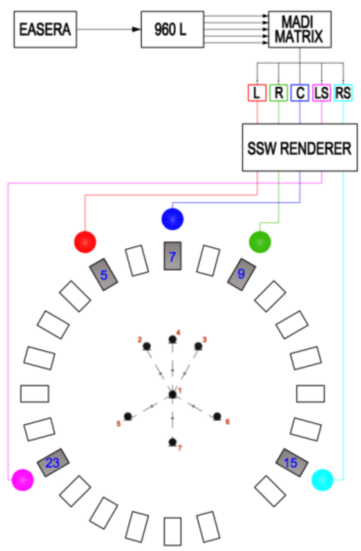

Figure 2: Loudspeaker array with chosen microphone positions for RIR measurements of the object-based SSW system. The virtual source positions are marked by colored circles.

4 Results and Discussion

The condensed results in the following section are part of the overall investigations.

4.1 System evaluation

Prior to detailed investigations on the influence of the Lexicon 960L reverberation unit, the overall system should be evaluated. Since the underlying principles of the two systems under test – 5.0 and SSW – are completely different, we expect different system behavior. To clarify this expectation, room impulse response measurements were conducted while bypassing the reverberation unit. Comparison were undertaken for the center position, i.e. center speaker for 5.0 and center virtual sound for SSW. Figure 3 shows the RIR of the respective 5..0 system and Figure 4 shows the rectified RIR from the SSW system.

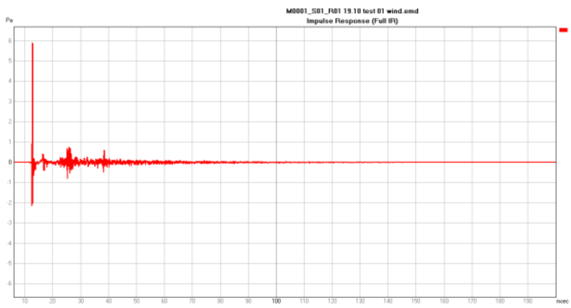

Figure 3: RIR of test 1 from 5.0 setup

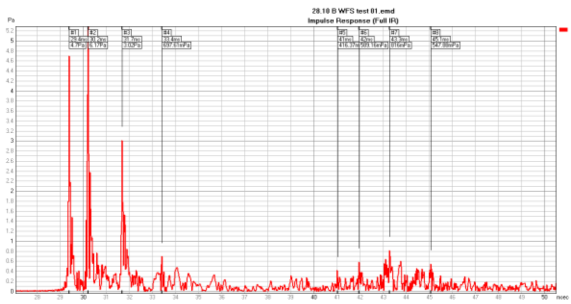

Figure 4: Rectified RIR of test 1 from SSW setup

The rectified RIR from the SSW setup shows very different temporal characteristics when compared to the RIR from the 5.0 setup. In particular, several closely spaced discrete components are visible after the arrival of the first wave front.

These discrete components are the result of spatial aliasing effects due to the loudspeaker setup as explained in [5], [11] and [12]. Since the loudspeakers that are activated by the SSW renderer – depending on the positioning of the point source behind the array – re-produce the virtual spherical wave front with different gain and delay settings, each loudspeaker re-produces a contribution of the wave front. As a result, the wave front is not perfectly superimposed at the microphone position, which is evident by these individual contributions. The first impulse (first arrival) is assumed to be from the centre speaker (due to the direct sound travelling the shortest path from source to receiver); the second impulse is a composite of the two adjacent speakers at a 15° opening angle while the third impulse is also a composite from the furthest two loudspeakers from the centre loudspeaker, at an equivalent 30° angle.

The spectral deviations between the systems are shown in Figure 5. The temporal disrecte components of the SSW system become noticeable by comb filtering effects for higher frequencies.

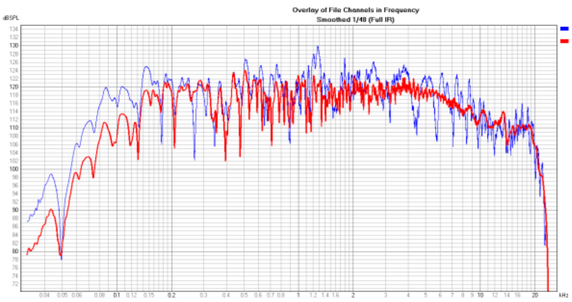

Figure 5: Frequency response curves of test 1 from 5.0 setup (red) and SSW setup (blue).

4.2 Temporal analysis

The reverberation unit's temporal influence on the RIR measurements is exemplarily analyzed in this section. For that matter test no. 13 (cf. Table 1) with increased early delay level is evaluated for both systems. The resulting effect can be determined by higher energy in the early part of the RIR which can be confirmed by observing the Schroeder backward integral of the 5.0 system in Figure 6 compared to the bypassed measurement in Figure 7.

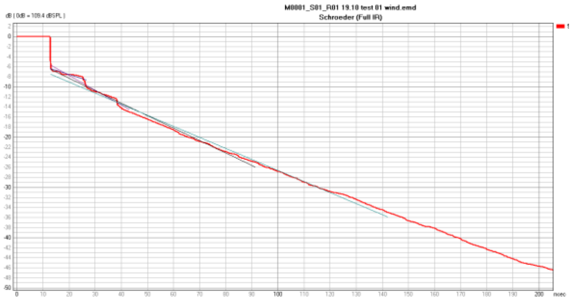

Figure 6: Schroeder curve of test 1 from 5.0 setup

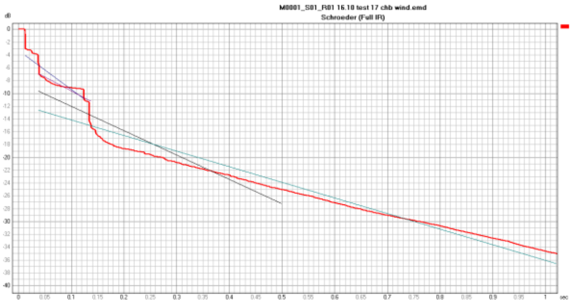

Figure 7: Schroeder curve of test 13 from 5.0 setup

The temporal differences between the two reproduction systems are shown in Figure 8. We can detect several delayed reflections for the SSW system due to spatial aliasing effects. However the overall slope of the Envelope Time Curve (ETC) shows obvious similarities between the systems.

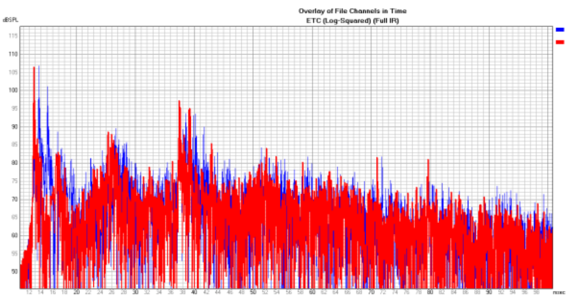

Figure 8: ETC for test 13 from 5.0 setup (red) and SSW setup (blue).

4.3 Quantitative analysis of RIRs

After exemplarily observing the temporal beahavior of the RIRs , this section will look at the results derived from the acoustical metrics. Results of test 4 and 13 (cf. Table 1) will be presented that were measured at the seven different microphone positions (as shown in Figures 1 and 2).

4.3.1 Test 4 Results

The first figure shows a graph with EDT values from the two setups for test four (Figure 9), followed by DRR (Figure 10) and T30 values (Figure 11) respectively. Blue indicates results from the channel-based 5.0 setup and orange shows results from the SSW setup.

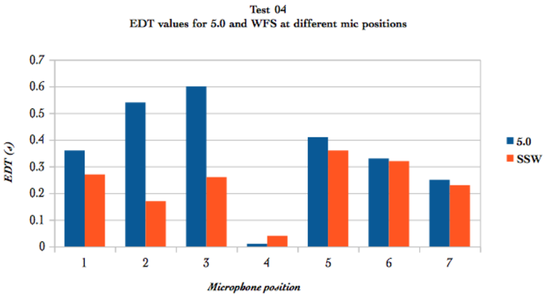

Figure 9: EDT values for test 4 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

EDT values are consistently higher in the 5.0 setup except in position 4, while the values from the back positions are relatively similar in both setups. Furthermore, the EDT patterns are relatively similar in both setups. The low EDT values from the SSW setup in positions 2 and 3 are assumed to be due to the high DRR values in the same positions as seen in Figure 18.

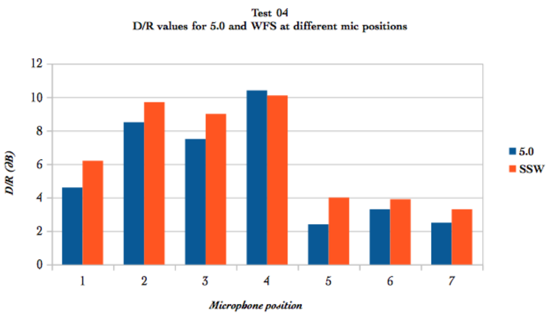

Figure 10: DDR values for test 4 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

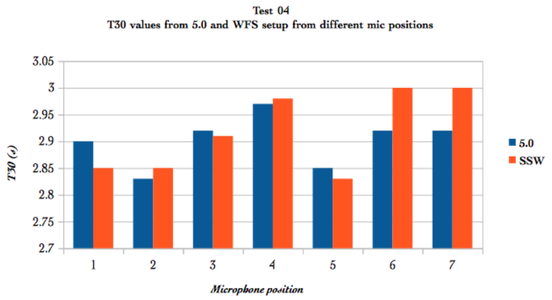

Figure 11: T30 values for test 4 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

In the case of T30 calculations, it is well known that values do not usually change between different positions within a room (assuming a diffuse sound field), in contrast to EDT which can change from place to place since it depends on temporal and directional distribution of the early reflections [10] [13]. However, in the measured RIRs, T30 values exhibit some fluctuations and such changes are relatively similar in both setups, with the SSW setup showing slightly higher values (approximately 0.10s higher) at positions six and seven suggesting a slightly higher perception of reverberation at the back of the loudspeaker array.

4.3.2 Test 13 Results

Results from test 13 showed similar behavior to the previous test (and the other tests), i.e., EDT was overall lower in the SSW setup (Figure 12), whereas D/R was higher (Figure 13). However, with the D/R ratio, the values extracted from the back microphone positions (i.e., positions 5, 6 and 7) appear to be fixed, suggesting a more consistent and possible diffuse sound field in the SSW setup. One

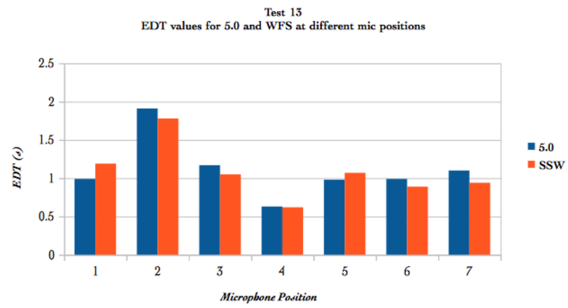

Figure 12: EDT values for test 13 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

explanation to this might be the result of the additional contributions present in the SSW RIRs which average out/smoothen the reflections present in the RIR.

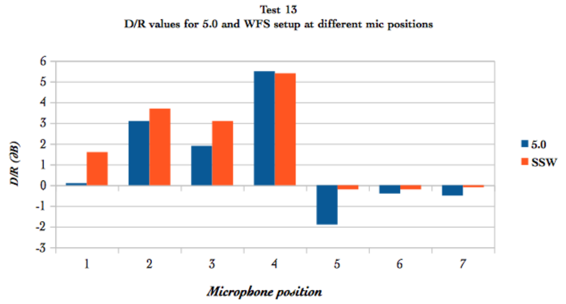

Figure 13: DRR values for test 13 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

T30 results shown in Figure 14 once again suggest very minor changes between the two setups suggesting that the perception of (the late part of) reverberation remains similar in both setups.

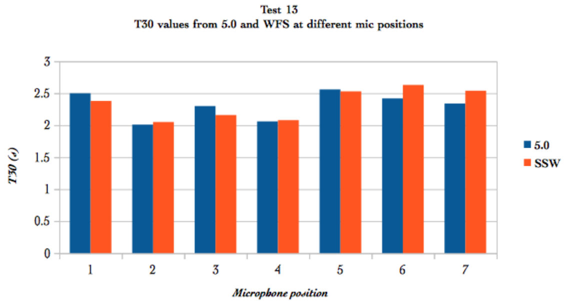

Figure 14: T30 values for test 13 from 5.0 setup (blue) and SSW setup (orange) at the 7 different mic positions.

4.4 Overall Findings and Results

After examining a couple of tests in detail and at different microphone positions, the results from the 20 tests at the centre of the loudspeaker array will be presented for a general overview of the results

EDT values from both setups appear to be very similar in both cases (Figure 15). EDT values were also analysed with the direct signal windowed out to examine the behaviour without any influence of the source signal on the RIR (as suggested by IEC) (Figure 16). In both scenarios results were similar within the two setups.

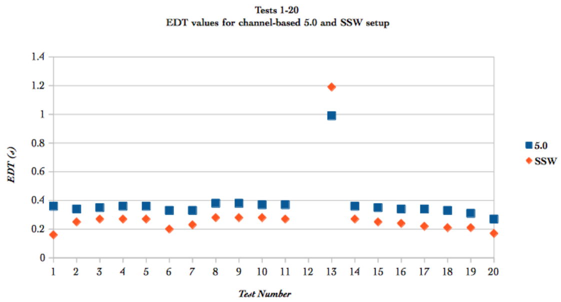

Figure 15: EDT results for all 20 measurements at mic position 1 for 5.0 setup (blue) and SSW setup (orange).

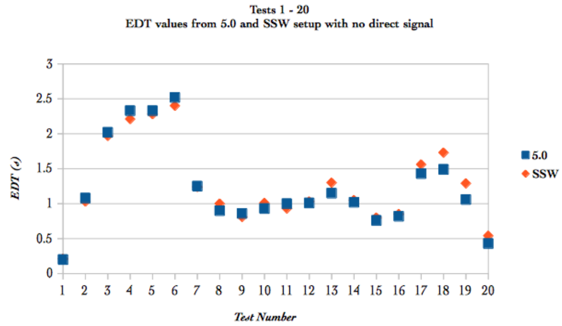

Figure 16: EDT results without direct signal for all 20 measurements at mic position 1 for 5.0 setup (blue) and SSW setup (orange).

The RT from the 960L from test 3 onwards was set at 2.363s. The following scatter plot in Figure 17 shows that the results in the SSW RIRs do not have an influence on T30 values and both follow very similar patterns according to the set parameters on the 960L. Furthermore, T30 values produced minor changes after windowing out the direct signal and this is expected since the T30 decay calculation is independent of the direct signal.

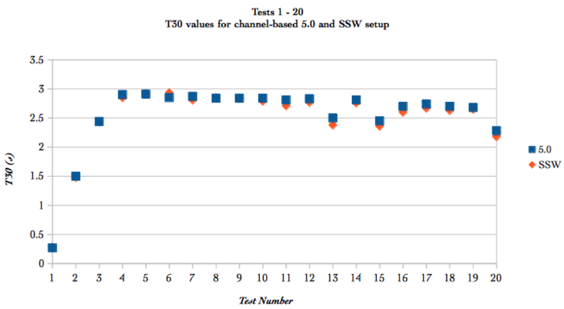

Figure 17: T30 results for all 20 measurements at mic position 1 for 5.0 setup (blue) and SSW setup (orange).

From the scatter plot in Figure 18 it can be seen that the DRR is always higher in the RIRs from the SSW setup. This suggests that the direct sound in the SSW setup appears to be stronger, and this is assumed to be due to the fact that the RIRs carry several impulses in the very early part due to the temporal artefacts produced by the SSW system. However, given that the reverberation parts also have additional components from this effect, the ratio of the sound pressure between the early and late parts should be relatively the same. In both setups, the DRR decreases from test 1 to test 2, which is expected due to the introduction of reverberation from the 960L in test 2.

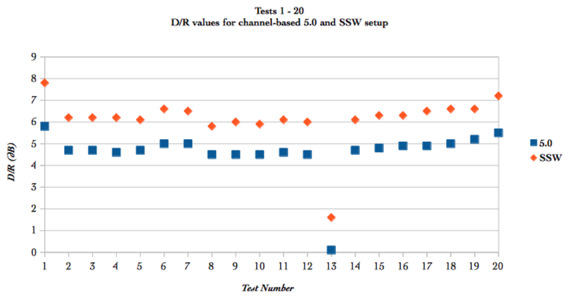

Figure 18: DRR results for all 20 measurements at mic position 1 for 5.0 setup (blue) and SSW setup (orange).

4.4.1 *Ltotal*

One rather curious result is from the *Ltotal* metric (which integrates the overall energy in the IR [14]; in this case the overall energy is presented in dBFS.). The results from the SSW setup have a constantly higher *Ltotal* value suggesting higher energy levels with the SSW loudspeaker array (Figure 19).

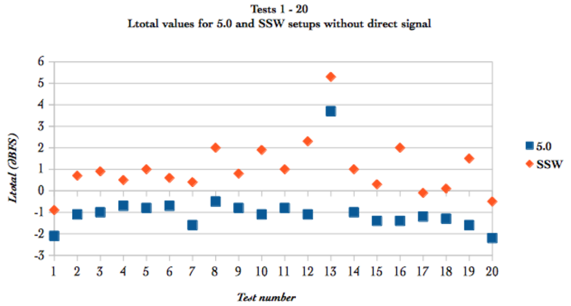

Figure 19: Ltotal results for all 20 measurements at mic position 1 for 5.0 setup (blue) and SSW setup (orange).

5 Conclusions

The conclusions which can be drawn from this research are summarized below:

(i) As hypothesized, the SSW setup used in this research skews the reflection patterns of the measured RIRs, primarily due to the spatial aliasing frequency of the setup, (which was estimated to be 390Hz). The result of these temporal artifacts conforms with the definitions of spatial aliasing by [6] [7]. The effects of spatial aliasing are also dependent on the virtual source positioning, the spectrum of the source and the listener/microphone position [7]. As a result, these time-shifted, discrete contributions cause spectral distortions and comb-filtering effects in the frequency responses from the SSW setup.

(ii) The different nature of the direct sound from the two setups due to spatial aliasing also changed the very early decay part of the Schroeder curve. As a result, the EDT values from the RIRs of the SSW were always lower (approximately 10ms) than those from the 5.0 setup and the assumption was made that the effects of the spatial aliasing frequency had some influence on the EDT. However, when the direct sound was windowed out from the RIRs, the EDT values produced very similar results which therefore opposed the previous assumption.

(iii) The overall decay characteristics of the Schroeder curve between both setups is very similar in all measurements. As a result, T30 values from the two setups were fairly similar and it is concluded that the SSW setup does not alter the overall reverberation time.

(iv) The other calculated acoustical metrics done in the research, i.e., C80, STI and BR also remained practically identical in both setups, suggesting that clarity does not suffer in the SSW setup.

References

[1] Senior, Mike: “Mixing Secrets for the Small Studio”. Focal Press, UK/USA, 2011.

[2] Rumsey, Francis: “Spatial Audio”. Focal Press, UK, 2001.

[3] Lund, Thomas: “Enhanced Localization in 5.1 Production”. AES 109th Convention, USA, September, 2000.

[4] Corey, Jason et al.: “Localisation of lateral phantom images in a 5-channel system with and without simulated early reflections.” AES Convention Paper 5673, Los Angeles, USA, October 2002.

[5] Verheijen, Edwin: “Sound Reproduction by Wave Field Synthesis”. PhD Thesis, TU Delft, 1997.

[6] Lembke, S et al: “Experimenting with an Object-based Sound Reproduction System for Professional Popular Music Production”. VDT International Convention - 28th Tonmeistertagung, November, 2014.

[7] Melchior, Frank: “Spatial Sound Design based on Measured Room Impulse Response”. PhD Thesis, Delft University and Fraunhofer IDMT.

[8] Huslebos, Edo: “Auralization using Wave Field Synthesis”. PhD thesis, Delft University, 2004.

[9] Farina, Angelo: “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique”. AES pre-print, presented at the 108th Convention, Paris, February, 2000.

[10] Kuttruff, Heinrich: “Room Acoustics”, 5th edition. Spon Press; Taylor & Francis, UK/USA, 2009.

[11] Berkhout, A.J. et al.: “Acoustic control by wave field synthesis”. J. Acoust. Soc. Am. 93, 2765–2778, 1993.

[12] De Bruijn, Werner Paulus Josephus: “Application of Wave Field Synthesis in Video conferencing”. PhD Thesis, TU Delft, 2004.

[13] Möser, Michael and Müller, Gerhard: “Handbook of Engineering Acoustics”. Springer, UK/USA, 2013.

[14] AFMG EASERA: “EASERA Tutorial 1.2”. AFMG Technologies GmbH, Berlin, 2012.