Audio Engineering Society

Convention Paper 9903

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

Dynamic diffuse signal processing for low-frequency spatial variance minimization across wide audience areas

Jonathan B. Moore and Adam J. Hill

Department of Electronics, Computing and Mathematics, University of Derby, Derby DE22 3AW, United Kingdom

Correspondence should be addressed to Jonathan Moore (j.b.moore@derby.ac.uk)

ABSTRACT

Diffuse signal processing (DiSP) is a method of decorrelating coherent audio signals which is applicable to various components of sound reinforcement systems. Previous tests have indicated that DiSP can successfully decorrelate multiple low-frequency sources, leading to the reduction of comb filtering effects. However, results also show that performance is variable with source material, and that effectiveness is reduced in closed acoustic spaces. In this work, a dynamic variant of DiSP is examined, where the decorrelation algorithm varies over time. The effectiveness of the processing is analyzed, and compared to static DiSP and unprocessed systems. Results show that dynamic DiSP provides superior low-frequency spatial variance reduction to static DiSP, due to improved decorrelation between direct sounds and early reflections.

1 Introduction

Sound reinforcement systems are subject to position-dependent frequency responses due to coherent interference between multiple loudspeakers outputting the same audio signal [1]. At low frequencies, this leads to large variances in frequency response across an audience area. A potential solution is to partially decorrelate each discrete source to cause a diffuse summation of overlapping dispersion patterns, hence avoiding problematic comb-filtering effects [2].

Diffuse signal processing (DiSP) was first described in [3] as a means of decorrelating audio output of large diffuse mode loudspeakers (DLMs) using synthetic, temporally diffuse impulses (TDIs). TDIs consist of an initial impulse followed by a low level, random phase exponentially decaying noise tail, as shown in Figure 1.1.

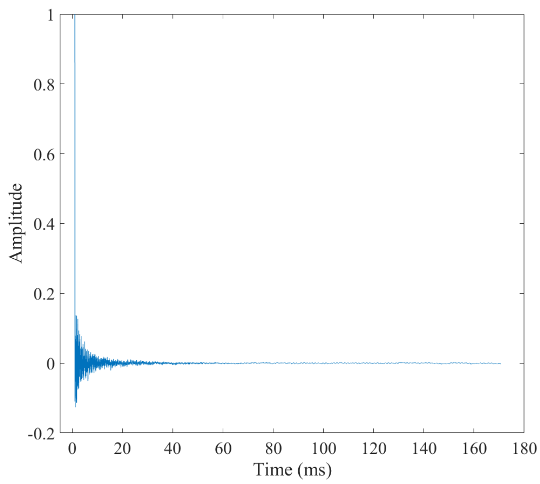

Figure 1.1 Time domain plot of an example TDI exhibiting initial impulse and decaying noise tail

To achieve decorrelation, each discrete source signal in the system must be convolved with a unique TDI.

In the initial description of DiSP [3], each TDI is time-invariant. For clarity, this form of processing will be henceforth described as *static DiSP*.

An investigation was previously conducted, looking into the use of static DiSP as a method of subwoofer decorrelation for the reduction low-frequency spatial variance in cinema sound [4]. Several methods of spatial variance minimization were assessed and compared, including multiple-point equalization [5, 6], chameleon subwoofer arrays [7] and static DiSP. Static DiSP was found to reduce spatial variance across audience areas, and whilst the reduction was not as great as that which was achieved with more calibration-intensive methods, it was noted that once the DiSP algorithm was optimized, it required no on-site or system-specific calibration; a significant advantage over other system optimization options.

The effectiveness of static DiSP for the reduction of spatial variance was further investigated in [8]. Several TDI generation methodologies described in [1] were examined, and the processing was assessed using multiple musical sources as well as a unit impulse and pink noise.

Results from an anechoic model found that low-frequency spatial variance (20-200 Hz) across a 10 m² audience area was reduced by 42%, averaged over all tested source material. However, results obtained in an image source model simulating a reverberant space show reduced spatial variance reduction after the application of static DiSP. It is concluded that if the TDIs remain unchanged over time, early reflections will maintain coherence with their direct sources, thus reducing the effectiveness of static DiSP [8].

This work investigates a potential solution to this issue, introducing a version of DiSP termed *dynamic DiSP* whereby TDIs are time-variant. The objective of dynamic DiSP is to partially decorrelate direct sources from each other as well as early reflections from direct sound (from the same source), thereby giving improved DiSP performance in closed acoustic spaces.

2 Diffuse signal processing overview

TDIs are synthesized by the summation of cosines of increasing frequency up to Nyquist. Each cosine has random phase and is subject to frequency dependent exponential decay. An all-pass frequency response is achieved via minimum phase equalization. In the time domain, the synthesized TDI is a single sample impulse followed by a low-level decaying noise tail [3]. A magnitude plot of an un-equalized TDI shows narrow spectral peaks and notches of random frequency throughout the spectrum (Figure 2.1).

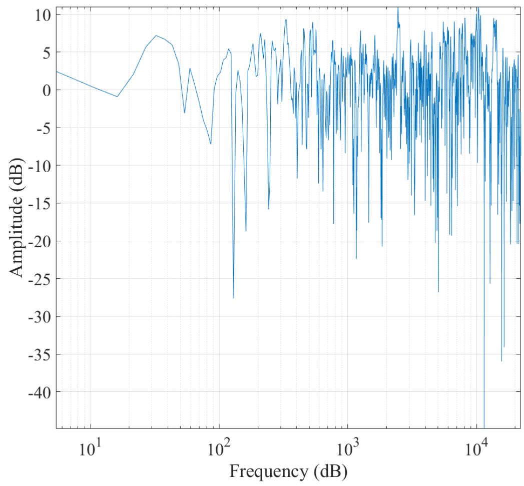

Figure 2.1 Unequalized magnitude response of an example TDI

Each TDI is generated with a different set of random phase values and decorrelation is achieved by using a different TDI for each system source. A full description of TDI generation is given in [3].

With the assessment of several TDI generation methodologies, the optimal TDI generation technique for the reduction of low frequency spatial variance was found to be a variable decay constant method using a uniform probability density function (PDF) for the phase values [8]. In this method, phase values are randomly generated with a uniform distribution between $\pm\pi$ with a constraining multiplier of 0.94. Exponential decay times for frequency components are selected at several frequency boundaries, with intermediate decay times obtained by linear interpolation between data points.

It is suggested in [8] that this method of defining decay time versus frequency allows for more flexible control of the compromise between audibility of the TDI and level of decorrelation achieved, as compared to methods described in [3]. In these methods, decay times are defined at the highest and lowest frequencies and linear or logarithmic interpolation is used to define intermediate frequency decays.

The variable decay constant method will be used in this work; however, imperceptibility of the filters is only achieved by careful selection of decay times for each frequency band. A subjective test to establish appropriate decay values is detailed in Section 4.

3 Static DiSP issues in acoustically-reflective environments

It was noted in [8] that static DiSP performance was reduced in a reflective acoustic model when compared to the anechoic model. This is illustrated by example in Figures 3.2, 3.3, 3.4 and 3.5.

Figure 3.2 shows the magnitude responses of 100 measurement positions evenly spaced across an audience area of 30 m x 30 m. Four point-source virtual loudspeakers were used, positioned at (3 m, 3 m), (7 m, 3 m), (23 m, 3 m) and (27 m, 3 m). The sound source was a unit impulse. A 2D image source model [9] was used as in [8], where the absorption coefficient for all surfaces was set to 1 to model an anechoic environment. Comb-filtering and large variances in magnitude response can be seen in the unprocessed responses in Figure 3.2.

Figure 3.3 shows the magnitude responses after application of static DiSP. The responses are more uniform, equating to greater consistency in listening experience over the audience.

Figure 3.4 shows the same simulation configuration, but with each surface in the model given an absorption coefficient of 0.2 to simulate a reverberant space. Figure 3.5 shows the magnitude responses of this simulation with static DiSP processing, using the same TDIs as in Figure 3.3. It can be seen that less uniformity in the magnitude responses is achieved.

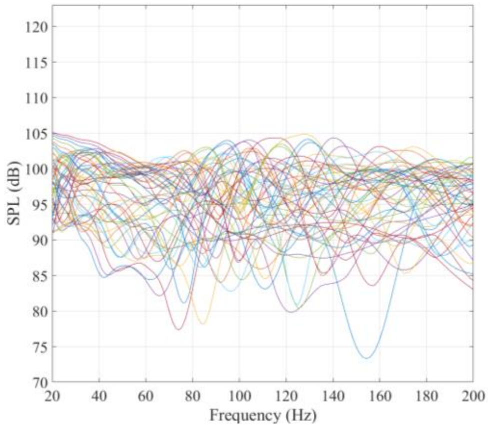

Figure 3.2 100 magnitude responses across 30 m x 30 m audience in anechoic model (no DiSP)

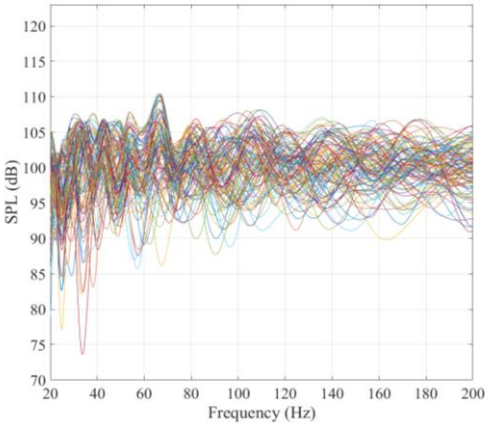

Figure 3.3 100 magnitude responses across 30 m x 30 m audience in anechoic model (w/static DiSP)

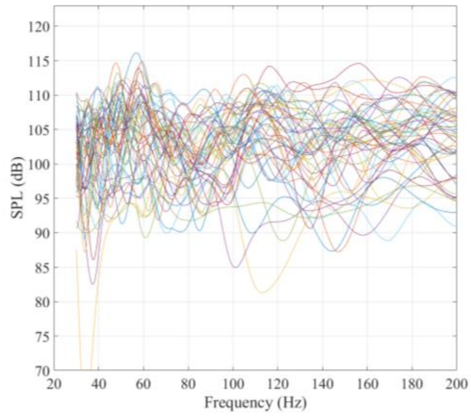

Figure 3.4 100 magnitude responses across 30 m x 30 m audience in reverberant model (no DiSP)

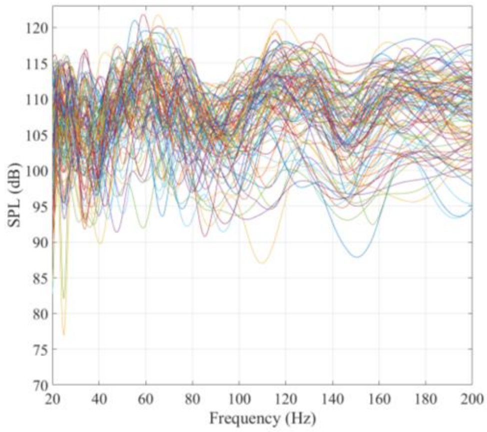

Figure 3.5 100 magnitude responses across 30 m x 30 m audience in reverberant model (w/static DiSP)

To quantify the differences in performance across these trials, the metric of spatial variance (*SV*) was used [10]. *SV* is derived here using Equation 3.1:

$$SV = \frac{1}{N-1}\sum_{i=1}^{N}\left|A_i - \overline{A_i}\right|^2 \qquad \text{(Eq. 3.1)}$$

where N represents the number of measurement positions, A_i represents the sound pressure level (SPL, dB) of a frequency bin measured at measurement position i, and $\overline{A_i}$ represents the mean SPL (dB) of that frequency bin at all measurement positions. *SV* is analysed here from 20 – 200 Hz and is given as a mean value over all frequencies.

Static DiSP reduced *SV* in the anechoic and reverberant models by 38.7% and 18.6%, respectively. The same TDIs were 20.1% less effective in an acoustically-reflective environment, indicating static DiSP isn't a robust solution.

4 Dynamic DiSP overview

Dynamic DiSP makes use of a large pre-generated library of TDIs. These TDIs are updated for every new audio frame for each loudspeaker in the system. By utilizing a pre-generated library of TDIs, the performance of the algorithm is not defined by a specific TDI/source material combination. Further to this, the rapid updating means that surface reflections can be decorrelated from direct sound, thus improving on static DiSP performance in reflective acoustic spaces. The effectiveness of the processing in this regard depends on many variables including room size and surface acoustic properties.

4.1 Perceptual artifact minimization

One of the primary goals of DiSP is perceptual transparency. Dynamic DiSP presents challenges in this regard. A key aspect is that the TDI library generation algorithm should reliably generate TDIs that are sufficiently decorrelated, but also perceptually transparent.

As TDIs are all-pass in nature, the primary subjective artefact of TDI convolution is the audibility of the noise tail that follows the initial impulse. This is related to the amplitude and duration of the noise tail, which is determined by the change of decay time over frequency. Longer decay times result in a noise tail with greater amplitude in comparison to the initial impulse, which has the effect of increasing the level of decorrelation achieved, but also the audibility of the TDI.

There is, therefore, a compromise between decorrelation and perceptible artifacts. This must be carefully balanced via careful selection of frequency dependent decay times within each TDI. It is suggested in [8] that decay times be defined at several intermediate frequency boundaries as well as at the maximum and minimum frequencies, with linear interpolation between data points. If correctly applied, this results in an ideal compromise, resulting in adequate inter-channel decorrelation without adverse perceptual artifacts.

4.1.1 Subjective test description

A subjective test was developed to establish suitable decay time limits for the following frequency bands: <63 Hz, 63-94 Hz, 94-125 Hz, 125-187.5 Hz, 187.5-250 Hz, 250-500 Hz, 500-1000 Hz, 1000-2000 Hz, 2000-4000 Hz and >4000 Hz.

The nature of the audible effects of TDI convolution are such that the filters are more audible with highly transient material. Therefore, two drum loops with sparse musical elements were chosen as source material, as they were found to be particularly revealing. The test was split into two components – one for each musical sample, to avoid listener fatigue, presented in a random order.

A real-time audio stream was used, enabling subjects to alter the TDI decay time for the frequency band under test via a slider in a GUI developed in MATLAB (Figure 4.1). A TDI was selected from a pre-generated library based on the slider's current value.

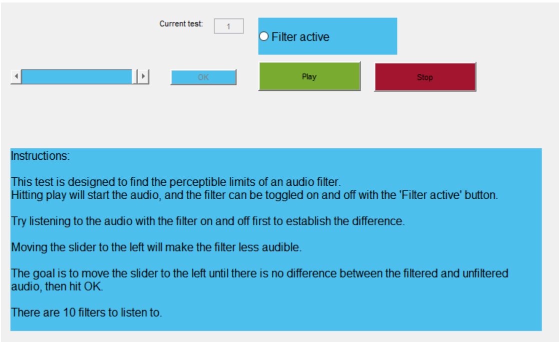

Figure 4.1 Test GUI presented to each subject

The test was performed over a pair of Sony MDR-V6 headphones [11] using a laptop running Windows 10. The GUI consisted of a play and stop button, a slider with which to alter the TDI decay time for the frequency band under test, a toggle switch with which to turn the filter on and off for direct comparison to unprocessed audio and a button to advance to the next band, once the audible threshold had been found.

Thirteen subjects aged 24-35 took the test (1 female, 12 male). Seven of these were considered to be experienced listeners. All subjects had normal healthy hearing.

4.1.2 Subjective Test TDI Library Generation

The test was performed with TDIs that only targeted decorrelation in one frequency band at a time (as detailed in Table 4.1), thus constraining perceptual artifacts to a specific frequency range. In this way, the audible limit of decay times vs. frequency band were judged in isolation.

Band (Hz)	Lower decay time limit (ms)	Upper decay time limit (ms)
<63	15.9	952.4
63-94	10.6	638.3
94-125	8	480
125-187.5	5.3	320
187.5-250	4	240
250-500	2	120
500-1000	1	60
1000-2000	0.5	45
2000-4000	0.25	40
>4000	0.04	8.3

Table 4.1 Upper and lower decay time limits vs. frequency band for the subjective test

Lower decay limits for each band were determined by the time taken for one cycle of the highest frequency to complete, while upper limits were determined by the time for sixty cycles of the highest frequency to complete, for all bands up to 1000-2000 Hz. The upper time limit for the band 1000-2000 Hz was based on ninety cycles, for 2000-4000 Hz 160 cycles, and for 4000-24000 Hz 200 cycles. The purpose of this was to generate TDIs that were extremely audible at the upper limit, and completely inaudible at the lower limit.

The slider was initially set to the maximum decay time for each band, providing clear audibility of the TDI convolution. The subjects were then asked to adjust the slider until they could hear no difference between the processed and unprocessed audio. This was repeated for each frequency band (which were presented in a randomized order).

4.1.3 Subjective Test Results

Figure 4.1 gives the results of the subjective test over all frequency bands and musical samples. Table 4.2 shows the median results for the threshold of audibility of decay time versus frequency band. Median values were used as opposed to mean due to non-normal distribution of the data.

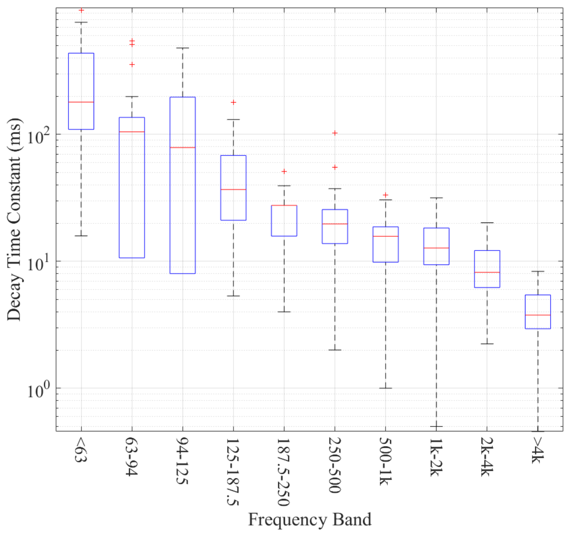

Figure 4.1 Logarithmic box and whisker plot of subjective test results for audible threshold of decay time vs. frequency band

Frequency band (Hz)	Median decay time audible threshold (ms)
<63	179.8
63-94	104.8
94-125	78.8
125-187.5	36.8
187.5-250	27.6
250-500	19.7
500-1000	15.7
1000-2000	12.7
2000-4000	8.2
>4000	3.7

Table 4.2 Median audible threshold of decay time vs. frequency band. The central frequency of each band is used to inform frequency decay time for dynamic TDI library generation

To apply the subjective test results to TDI library generation, the variable decay constant method was used, with decay times defined at the central frequency of each band in the subjective test. Linear interpolation was used for intermediate frequencies. This leaves the question of decay time at Nyquist. As decorrelation is not desired in this application from 4 kHz to Nyquist, decay times above 4 kHz were set to 0.1 ms to minimize any perceptual effects at high frequencies. It should be noted that the subjective test used static TDIs, and so whilst these results should reliably deliver perceptually transparent TDIs for use with static DiSP, further subjective tests may be required to ensure that this data is compatible with dynamic DiSP.

5 Dynamic DiSP algorithm

For optimal functionality, a TDI library should be large enough so that TDIs are only repeated once early reflections are fully decorrelated. The required size of the library will therefore depend on the dimensions of the space, the acoustic properties of surface materials and the number and positions of sources in the system. In this case, a “more is better” approach is desirable, as long as there is sufficient memory in the processing unit to store the library.

For this work, a library of 100 TDI sets of four, with a TDI length of 250 ms was used (one TDI for each loudspeaker per set). A frame overlap of 75% was used to ensure seamless audio output, which corresponds to the library repeating itself once every 6.25 seconds. TDI selection for each time frame was made by cycling through the library. A randomized selection method is also possible, however this increases complexity as checking for TDI repetition must be included.

The current audio frame is convolved with the TDI set, generating N decorrelated audio channels (for a system with N loudspeakers) which are then outputted via the sliding window.

5.1 Selection of TDI update rate

Size of the TDI processing frame (which dictates TDI update rate) is key. Too short and adequate decorrelation will not be achieved due to loss of frequency resolution. Too long and processing time (especially when a large number of sources are to be processed) may result in latency in the audio stream. More importantly, too large a processing frame could cause the direct sound and early reflections to be processed by the same TDI, thus negating the purpose of dynamic DiSP. Correct frame size selection should result in a TDI update rate which is faster than the first reflection arrival time. This will therefore depend on the topology of the acoustic space the system is operating within.

Figure 5.1 shows example required TDI update rates for various room dimension ratios (length : width : height). To establish the required TDI update rate, the path length difference between direct sources and 1st order reflections is calculated. In the present application only low frequency decorrelation is desired, so only paths leading to a delay in arrival time equal to the half period of 200 Hz and greater are considered. It should be noted that more complex, real world room topologies will require a more detailed analysis.

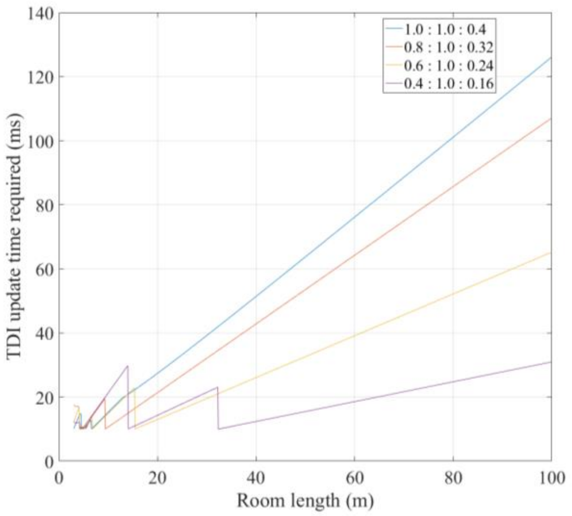

Figure. 5.1. Required TDI update times for 4 room size ratios (length : width : height) based on difference in arrival times for direct sounds and first order reflections at a central measurement location

6 Simulation configuration and results

An assessment of dynamic DiSP versus static DiSP for the reduction of low frequency spatial variance across audience areas was performed. The real-time processing approach for both algorithms was identical, apart from the TDI set being changed for every audio frame with dynamic DiSP. For static DiSP, the TDI set remained fixed.

An image source model was used, with a simulated room size of 30 m x 30 m. Surfaces were given absorption coefficients of 0.29 (walls of plasterboard), 0.45 (porous ceiling tiles) and 0.6 (to simulate typical audience absorption). Four point-source virtual loudspeakers were positioned at (3 m, 3 m), (7 m, 3 m), (23 m, 3 m) and (27 m, 3 m). Reflections up to 15th order were simulated. A central measurement position resulted in a first reflection arrival 40 ms after the direct sound. Therefore, a TDI update rate of 40 ms was chosen for dynamic DiSP. In order to compare anechoic and reflective acoustic performance, tests were performed with an identical model configuration apart from all surface having the absorption coefficient of 1.0 for the anechoic model.

To quantify performance, spatial variance (*SV*) was calculated using Eq. 3.1 for 100 measurement positions across the audience area, averaged over a frequency range of 20 - 200 Hz. A reduction in *SV* after the processing is indicative of a more uniform low frequency magnitude response across the audience area, which is the aim of DiSP.

To obtain the presented results, a musical sample was used as a source signal and the transfer function of each measurement position was derived by dividing the measurement position's measured complex frequency response by the input signal's complex frequency response. In order to analyze performance over different time frames, the resulting measurement position transfer functions were converted back to impulse responses via an inverse Fourier transform and windowed accordingly, before conversion to magnitude responses (with a fixed 8192 resolution) for the *SV* calculation.

Analysis was performed over several time frames (50 ms, 100 ms, 170 ms, 250 ms, 500 ms, 1.0 s, 5.0 s, 10 s). The 170 ms analysis window is of particular interest as it's in line with the auditory temporal integration time as suggested in [12] and [13].

Figures 6.1 and 6.2 give the results with static and dynamic DiSP in the anechoic and non-anechoic models, respectively. In both models, dynamic DiSP is shown to reduce spatial variance by a greater amount than static DiSP. In the reflective model, static DiSP performance was reduced as expected due to the correlation of early reflections with their direct sources leading to comb-filtering. Dynamic DiSP performance, however, is unaffected due to the successful decorrelation of early reflections from direct sources. This leads to greater diffusion in the overall sound field, and a more even distribution of sound energy.

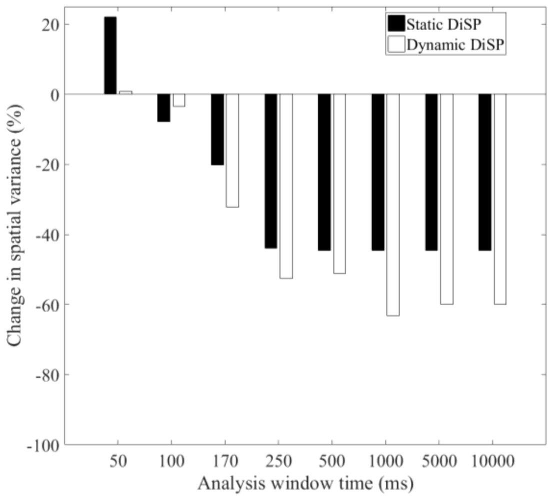

Figure 6.1 Change in spatial variance (20 – 200 Hz) for the anechoic model

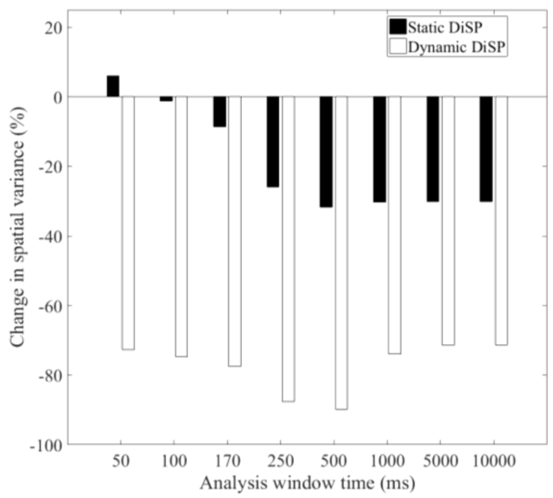

Figure 6.2 Change in spatial variance (20 – 200 Hz) for the reflective model

In the anechoic model, it can be seen that both the static and dynamic algorithms struggle to provide spatial variance reduction when analyzed in windows shorter than 250 ms. This is due to the fact that the TDIs are 250 ms in length, so the full DiSP effect can't take hold in an anechoic environment before a full TDI has completed. In the reflective model, this isn't an issue for dynamic DiSP as early reflections arrive prior to the first TDI completion, thus providing the necessary decorrelation.

Dynamic DiSP performance increases with analysis window length in the reflective model up to 500 ms. This appears to indicate that higher-order reflections are more prevalent for analysis windows above 500 ms. These reflections will be highly-correlated since the image source model doesn't incorporate frequency-dependent absorption. In real-world scenarios, these reflections will be naturally decorrelated from the direct sound and early reflections.

Critically, in the reflective model spatial variance is reduced within the auditory temporal integration time (170 ms), thus giving indication that the effect will be perceptually significant in terms of reduction in magnitude response variance across an audience area.

6.1 Dynamic DiSP with a single loudspeaker

An interesting scenario is that of the performance of dynamic DiSP with a single loudspeaker. It is expected that constantly updating TDIs will decorrelate early reflections from each other as well as the direct sound from itself (over time). Therefore, it is possible that there will be an improvement in low frequency magnitude response consistency over an audience area, even with a single source.

Figure 6.3 shows *SV* reduction using static DiSP and dynamic DiSP for a single source. The image source configuration was identical to that used previously, with a single virtual loudspeaker placed centrally, 0.4 meters away from the front wall. Anechoic results are not presented as no spatial variance is possible with a single source in a free-field.

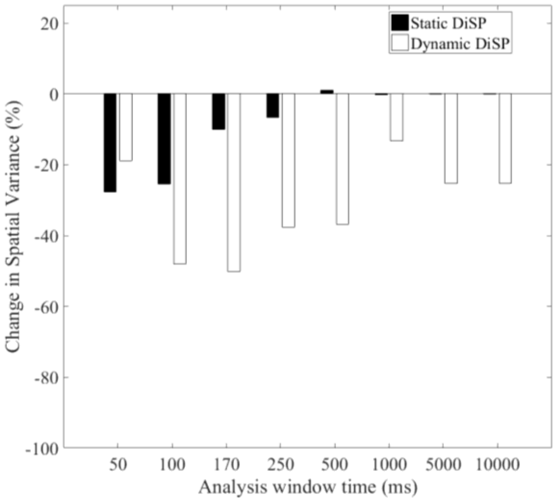

Figure 6.3 Change in spatial variance (20 – 200 Hz) with a single loudspeaker for the reflective model

It can be seen that dynamic DiSP is effective in the reduction of spatial variance in a reflective acoustic model with a single source. For short analysis windows (< 250 ms) static DiSP can also be seen to have an effect. This is due to the TDI length used of around 250 ms (note that this is not the length that the noise tail has significant energy, but the overall length of the decorrelation impulse). Until sufficient time has elapsed for a full TDI length, there is still variation in successive audio frames. Once this time has elapsed, static DiSP is ineffective as successive audio frame correlation is maximized.

This issue is negated with dynamic DiSP, whereby each audio frame is decorrelated from the next, with an overlap of 75% utilized. As can be seen from the results in Figure 6.3, dynamic DiSP provides upwards of 50% reduction in spatial variance even with a single physical degree of freedom (the loudspeaker).

These results are encouraging, although experimental investigations are required to verify the effectiveness of the processing and that perceptual transparency is maintained.

7 Conclusions

The results presented in this paper suggest that dynamic DiSP may be a useful tool in the minimization of low frequency spatial variance across audience areas. It has been shown that static DiSP is less effective in reflective environments, due to lack of decorrelation between the direct sound and early reflections.

A subjective test to establish frequency-dependent decay time limits for use in TDI generation has been described, the results of which can be used for the generation of perceptually-transparent TDIs. However, as the subjective test was performed with a static algorithm, further testing needs to be carried out to ensure these decays times are suitable with dynamic DiSP.

Rapidly changing TDIs over time makes the processing more audible, because slight timbral changes are evident across different TDIs. When the TDIs remain constant, this effect isn’t present. It is suggested that for perceptually-transparent dynamic DiSP, care must be taken to either fade or interpolate between successive TDIs to provide a smooth transition as opposed to an abrupt change.

As the model used was a simple 3-dimensional rectangle which did not simulate frequency-dependent absorption, differing behavior may be observed with more complex real-world acoustic environments.

The primary goal of dynamic DiSP is that it should be a turn-key solution, with limited system calibration required. This would make dynamic DiSP much simpler to implement in practice as compared to other calibration-intensive techniques.

The simulated results given in this work provide an encouraging first step towards implementing dynamic DiSP into real-world systems. The focus now is to begin experimentation in a controlled laboratory environment and, eventually, progress to real-world in-situ testing with large-scale sound reinforcement systems.

References

[1] Eargle, J. and Foreman, C.; "Audio engineering for sound reinforcement", *Hal Leonard Corporation.* 2002

[2] Hill, Adam. J.; Hawksford, Malcolm, O. J. "On the perceptual advantage of stereo subwoofer systems in live sound reinforcement," *135th Convention of the Audio Engineering Society, convention paper 8970*. October, 2013

[3] Hawksford, M.O.J.; Harris, N. "Diffuse signal processing and acoustic source characterization for applications in synthetic loudspeaker arrays," *112th Convention of the Audio Engineering Society, convention paper 5612*. May, 2002.

[4] Hill, Adam. J.; Hawksford, Malcolm, O. J.; Newell, Philip. "Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas: Effective and Practical Alternatives to Current Sub-Optimal Calibration Strategies," *The 57th International Conference of the Audio Engineering Society*. March, 2015.

[5] Newell, P.; G. Leembruggen; K. Holland; J. Newell; S. Torres Guijarro; D. Gilfillan; D.S. Dominguez, S. Castro. "Does 1/3rd octave equalization improve the sound in a typical cinema?" *Proc. IOA – Reproduced Sound 2011, vol. 33, pt. 6*, 2011.

[6] Elliott, S.J.; P.A. Nelson. "Multiple-point equalization in a room using adaptive digital filters." *Journal of the Audio Engineering Society, vol. 37, no. 11, pp. 899-907*. November, 1989.

[7] Hill, A.J. "Analysis, modeling and wide-area spatiotemporal control of low-frequency sound reproduction.*" Ph.D. Thesis, University of Essex, UK*. January, 2012.

[8] Moore, J. B.; Hill, A. J. "*Optimization of Temporally Diffuse Impulses for Decorrelation of Multiple Discrete Loudspeakers*" The 142nd Convention of the Audio Engineering Society, convention paper 9794. May, 2017.

[9] Allen, J.B.; D. A. Berkley. "Image Method for Efficiently Simulating Small-Room Acoustics," *The Journal of the Acoustical Society of America, vol. 65, no. 4*, 1979.

[10] Welti, T.; Devantier, A. "Investigation of bonello criteria for use in small room acoustics." *JAES, Vol. 54, Issue 5, pp. 347-364.* Oct, 2009

[11] https://esupport.sony.com/US/p/model-home.pl?mdl=MDRV6

[12] Ross, B.; Picton, T.W.; and Pantev, C, "Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field." *Hearing research, 165(1), pp.68-84*, 2002.

[13] Efron, R., "The minimum duration of a perception." *Neuropsychologia, 8(1), pp.57-63,* 1970.