Audio Engineering Society

Convention Paper 10181

Presented at the 146th Convention
2019 March 20–23, Dublin, Ireland

Optimizing wide-area sound reproduction using a single subwoofer with dynamic signal decorrelation

Adam J. Hill and Jonathan B. Moore

School of Electronics, Computing & Mathematics, University of Derby, UK

Correspondence should be addressed to Adam Hill (a.hill@derby.ac.uk)

ABSTRACT

A central goal in small room sound reproduction is achieving consistent sound energy distribution across a wide listening area. This is especially difficult at low-frequencies where room-modes result in highly position-dependent listening experiences. While numerous techniques for multiple-degree-of-freedom systems exist and have proven to be highly effective, this work focuses on achieving position-independent low-frequency listening experiences with a single subwoofer. The negative effects due to room-modes and comb-filtering are mitigated by applying a time-varying decorrelation method known as dynamic diffuse signal processing. Results indicate that spatial variance in magnitude response can be significantly reduced, although there is a sharp trade-off between the algorithm's effectiveness and the resulting perceptual coloration of the audio signal.

1 Introduction

It is essential in the optimization of small room sound reproduction to achieve position-independent listening experiences. Although there is currently no reliable method of controlling the subjective impression of sound reproduction, care must be taken to at least deliver the same objective acoustic signal to each listener's ears to provide what some in the industry refer to as the "democracy of sound."

Above a room's modal frequency band (the frequency range where room-modes are perceptible and problematic in terms of reproduction consistency [1]) recent research has focused on the use of centrally-located horizontal loudspeaker arrays to achieve high-quality immersive audio for multiple listeners [2-6]. This is usually achieved with a combination of wave-field synthesis, interaural crosstalk cancellation and head-related transfer functions. Unfortunately, these systems are necessarily limited in the low-frequency range due to their relatively small size [7], therefore the benefits of such arrays do not extend to the subwoofer band. A separate solution is required.

There exist multiple techniques that achieve highly-uniform listening experiences in the modal region [8-22]. Nearly all of these approaches require multiple subwoofers (or at least a single subwoofer with multiple degrees of freedom) with dedicated signal processing and amplification. In addition to the extra hardware and processing, most of these approaches necessitate calibration measurements.

While such methods achieve useful results and are commonly implemented in high-end sound reproduction systems, they are impractical for the vast majority of consumers as most individuals aren't likely to be willing to purchase multiple subwoofers and take time-consuming measurements to optimize their sound system at home. For an optimization technique to have mass appeal and adoption, it should be implemented entirely within the system processor with no extra hardware or calibration measurements.

This paper details an investigation into small room low-frequency sound reproduction optimization using a single conventional subwoofer with *dynamic diffuse signal processing* (DiSP). Section 2 gives a brief overview of DiSP, with Section 3 detailing the adopted experimental procedure for this investigation. The results are presented and analyzed in Section 4, with the paper concluded in Section 5.

2 Diffuse signal processing

There are numerous signal decorrelation algorithms for a range of applications including sound reproduction [23] and reinforcement [24], echo-cancellation [25], pseudo-stereo synthesis [26], headphone externalization [27], control of apparent source width [28] and reverb synthesis [29]. For signal decorrelation in regards to small room low-frequency sound reproduction, the selected algorithm must achieve sufficient decorrelation down to 20 Hz without perceptually degrading audio quality.

An extensive review of the available decorrelation methods was recently carried out by one of the authors [30] with the conclusion that a modified version of DiSP, which was first developed for use with distributed mode loudspeakers (DMLs) [31], is the best possible solution for such applications.

DiSP operates by synthesizing what are known as temporally diffuse impulses (TDIs), first described in [31]. TDIs consist of a single sample impulse followed by an exponentially-decaying noise tail with frequency-dependent decay characteristics. The noise tail is synthesized by randomizing the impulse's phase response over frequency, where randomization is controlled by a probability density function (PDF) and the frequency-dependent decay is defined so as to minimize signal coloration [30].

Previous work has found that a uniform PDF provides the best possible DiSP performance and decay time constants have been determined through a series of formal listening tests [32]. A full explanation and analysis of the synthesis process is detailed in [30]. An example TDI is shown in Figure 2.1.

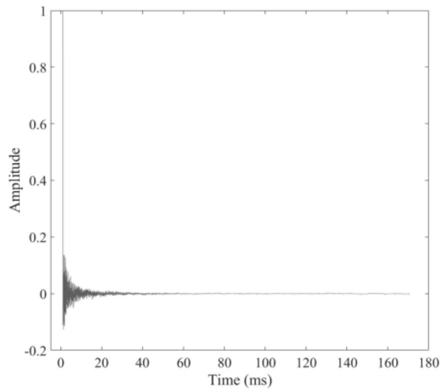

Figure 2.1 Example TDI for use within DiSP [32]

Synthesizing a single TDI for each degree of freedom within the system and convolving the TDIs with the input audio results in the perceptually-transparent decorrelation of each sound source in order to avoid effects due to coherent acoustic interference. This is particularly useful in large-scale sound reinforcement applications, providing significant reductions in seat-to-seat frequency response variance [32].

The problem when applying this sort of optimization in small closed spaces is that while direct sounds emitted from loudspeakers will be adequately decorrelated, early reflections will remain correlated with their corresponding direct sound, causing acoustical issues related to comb-filtering and room-modes. It was previously shown that applying *static DiSP* in small rooms results in minimal improvement in consistency over an audience area [32].

In order to mitigate the problems encountered by static DiSP in small rooms, a time-varying implementation of DiSP was developed, termed *dynamic DiSP* [33]. In this approach, a library of TDIs are generated for each sound source. A different TDI is selected at random for each audio frame and sound source with a controllable level of interpolation between adjacent TDIs to avoid audible coloration as the TDIs are updated (more on this in Section 3.2).

Dynamic DiSP results in the decorrelation of direct sound from each source as well as reflections within a listening space. While this won't entirely eliminate comb-filtering and room-modes, it serves to reduce these issues without negatively affecting sound quality [33].

An early informal listening test indicated that dynamic DiSP smeared transient responses, which resulted in low-impact listening experiences. To overcome this, a transient content detector adapted from [34] was built into the algorithm so that sharp signal transients bypass DiSP in order to maintain a good transient response. The effectiveness of this was verified in [33], where system optimization was only slightly less effective than for standard DiSP.

Overall, dynamic DiSP has been shown to reduce spatial variance in the low-frequency magnitude response across a listening area in a small room by nearly 60% in certain cases when using two subwoofers [32]. While DiSP is more effective with a larger number of sources, what happens with only one degree of freedom? This is the focus of this work.

3 Single source optimization

For a sound reproduction optimization method to have mass appeal it must require little to no extra effort or expenditure by users. Dynamic DiSP offers an interesting solution, whereby low-frequency spatial consistency can be achieved with a single conventional subwoofer (one degree of freedom) with no calibration measurements required.

This idea was first investigated using an image source model in [32] and later with real-world measurements in [24], with peak spatial variance reductions of approximately 50% and 25%, respectively. In both instances there was no attempt to optimize TDIs for the single-source application, as this wasn't the primary focus of either piece of work.

In the current research, two system configurations are explored in a typical living room environment: subwoofer only and full system (left, right, subwoofer). The inclusion of the full system configuration is relevant, considering that the subwoofer is unlikely to reproduce low-frequencies in isolation; it is common to have stereo loudspeakers which extend in frequency response to quite low-frequencies. Therefore, it is expected that there will be a frequency crossover region whereby all system elements are contributing to sound reproduction. In this case, DiSP can be applied to all loudspeakers over the crossover frequency region and below (providing two extra degrees of freedom).

While this isn't strictly a single source application of DiSP, it represents a typical sound reproduction system. All sound sources capable of low-frequency reproduction should be utilized to ensure a spatially-invariant listening experience. This recommendation was first proposed in the context of cinema B-chains in [35].

3.1 Experiment configuration

The experiment took place in a domestic living room of dimensions 4.85 m x 4.08 m x 2.30 m (width x depth x height), where the front and left walls were largely covered by bookcases, the right wall was wallpapered plaster and the rear wall was predominantly floor-to-ceiling glass windows surrounded by painted plaster. The ceiling was painted plaster and the floor was laminate with a 1.8 m x 1.2 m thick rug centered in the listening area.

Five listening locations were chosen, corresponding to seats on the two couches in the room. The precise listening locations are given in Table 3.1, where the room's origin was set to the front left floor corner. The system's stereo pair of loudspeakers were Audio Physic Classic Compacts [36] driven by a NAD D 3020 V2 amplifier [37] and the subwoofer was a Tannoy TS2.8 [38]. The precise locations of the loudspeakers are given in Table 3.2.

#	Width	Depth	Height
1	1.60 m	3.20 m	1.00 m
2	2.20 m	3.20 m	1.00 m
3	3.10 m	2.30 m	1.00 m
4	3.10 m	1.65 m	1.00 m
5	3.10 m	1.20 m	1.00 m

Table 3.1 Listening location coordinates

Loudspeaker	Width	Depth	Height
Left	1.25 m	0.35 m	1.25 m
Right	2.50 m	0.35 m	1.25 m
Subwoofer	0.95 m	0.20 m	1.75 m

Table 3.2 Loudspeaker location coordinates

The stereo speakers weren't crossed over, allowing them to extend to as low a frequency as they were capable of reproducing (published as down to 50 Hz [36]). The subwoofer's low-pass filter cutoff frequency was set to its maximum allowable value of

250 Hz. While it is unlikely that the subwoofer would operate to this high a frequency in practice, allowing the wider bandwidth helps to highlight DiSP performance at low and low-mid frequencies and also showcases the importance of applying DiSP to all loudspeakers.

3.2 TDI library generation

In order to determine optimal values for the TDI synthesis variables, a set of TDI libraries were generated. In all cases, a uniform PDF was used with frequency-dependent decay times set using the variable decay method, as described in [33].

The first synthesis variable explored was TDI duration. The longer TDIs give finer frequency resolution, which is essential for significant low-frequency decorrelation. Three TDI durations were investigated: 170 ms, 341 ms and 683 ms.

Of course, using long TDIs means more signal latency due to the increased amount of signal processing. Latency isn't addressed here, but further work should investigate partitioned convolution as a potential solution for low-latency real-time DiSP.

The second synthesis variable investigated was TDI update rate. This dictates how often the TDI applied to each sound source is changed. As highlighted in [32], TDIs must be updated fast enough to sufficiently decorrelate the direct sound from early low-order reflections. Otherwise, there will be little difference between dynamic and static DiSP performance. Given the room dimensions and loudspeaker locations in this particular experiment, the required update rate was calculated as just under 10 ms. In order to judge the sensitivity of this variable, update rates of 5 ms, 10 ms and 15 ms were investigated.

When TDI update rate is very small (such is the case here), abruptly switching TDIs results in an audible transition. To avoid this, a method of TDI interpolation was developed [30] whereby intermediate TDIs are calculated and inserted between the two adjacent TDIs in order to provide a smoother, less audible transition. The greater the so-called *interpolation factor* (number of intermediate TDIs) the less audible the transition (at the expense of DiSP performance). In this work, interpolation factors of 2, 10 and 30 were investigated.

TDI duration is considered within the TDI library generation process, while update rate and interpolation factor are applied during DiSP test signal generation (as discussed in Section 3.3).

The three chosen TDI durations resulted in three TDI libraries. Each library was generated at a sampling rate of 48 kHz with decorrelation applied to the entire frequency range up to Nyquist (0 Hz – 24 kHz). Each library was generated to contain 100 TDIs for use in the dynamic DiSP algorithm.

An additional three TDI libraries were generated for the full system configuration (left, right, subwoofer). These were required due to the two additional degrees of freedom, therefore requiring two extra sets of 100 TDIs. Each TDI pair in the library could exhibit no greater than 0.1 correlation. This ensures that at any point in time, the three loudspeakers' signals will be sufficiently decorrelated from each other.

3.3 DiSP test signal generation

The test signal used in this work was a 17^{th} order maximum length sequence (MLS) with a duration of 2.73 s at 48 kHz. This signal was repeated four times, resulting in an overall duration of just under 10 s.

The raw test signal was run through the DiSP algorithm for all possible combinations of TDI update rate and interpolation factor. This was repeated for the subwoofer-only and full system test configurations. Each configuration, therefore, had 27 test signals. With five listening locations, this necessitated 135 measurements for each configuration. In all cases, the raw MLS was first passed through a complementary 2^{nd} order Butterworth crossover at 250 Hz. DiSP was only applied to the sub-250 Hz band.

All test signals were saved to .wav format and a script in MATLAB was developed to automatically conduct measurements and store the resulting data as .wav files for analysis.

All measurements were taken with an Earthworks M30/BX measurement microphone [39] at a height of 1.00 m off the floor, which is the approximate ear height of a listener sitting on one of the couches. The microphone was fed into a Sound Devices USBPre [40], which was connected to a laptop running Windows 10 and MATLAB R2018a [41].

In addition to the MLS signal, two musical samples were processed with identical DiSP settings in order to judge perceptual effects of the processing. One sample was instrumental (*Cousin John* by Marcus Miller) and one was acapella vocals (*These Bones 'Gwine Rise Again* by The Blind Boys of Alabama). While not used directly for objective analysis, these samples were crucial to judge whether particular DiSP settings avoided audible signal coloration.

4 Results and analysis

Due to the number of measurements required in this work, it wouldn't be useful to view the data as a collection of frequency responses, as is typical of work focused on spatial variance minimization. Due to the time-varying nature of dynamic DiSP, frequency response analysis of the overall measurement isn't likely to reveal the true nature of the system's electroacoustic response. Instead, two forms of analysis were chosen. First, a cumulative analysis was performed, where a Hann window was applied beginning with a length of 50 ms and increasing by 50 ms until it spanned 5 s.

The data was analyzed for magnitude response spatial variance (using the standard deviation-based calculation [9], as opposed to the variance-based calculation [20]). All data was smoothed using equivalent rectangular bandwidths, according to [42], to ensure the analysis was perceptually accurate. Three data points were extracted to represent the short-term response (50 ms), the approximate integration time of the human hearing system at low-frequencies (270 ms) [43,44] and the long-term response (5000 ms). While static DiSP can provide good spatial variance reduction for the short-term response, it fails to reduce spatial variance over the mid- to long-term [32]. Dynamic DiSP should provide more consistent spatial variance reduction.

The second analysis was a partitioned approach, where a sliding 270 ms Hann window was used to analyze the DiSP spatial variance reduction over five seconds of the measurement. 270 ms was used as the window length as this corresponds to the hearing system's integration time at low-frequencies [43,44].

It was expected that spatial variance would fluctuate over time with this analysis, therefore three values were extracted from each measurement: minimum, mean and maximum spatial variance. This gives a good idea of how the system behaves over time.

Results from both analyses are given in Tables 4.1 & 4.2. Table 4.1 gives the results directly as spatial variance, with the unprocessed system's spatial variance values included. Table 4.2 presents the results in terms of percentage change in spatial variance from unprocessed to DiSP-processed.

Both forms of data presentation are necessary in this work. First, inspecting the spatial variance values directly allows the determination of their audibility. It has been found through recent listening tests [30] that in the 20 – 250 Hz range spatial variance below 1.38 dB is inaudible. This means that any system with less than 1.38 dB spatial variance can be considered fully optimized (i.e. the listening experience is consistent across all locations). Such situations are indicated by dark green in Table 4.1. As spatial variance rises, the highlighting color goes from green to yellow to red, red representing highly-audible spatial variance.

The percentage change data in Table 4.2 is also useful as it gives a clear indication as to how much a system has improved post-DiSP. Green indicates significant spatial variance reduction while orange and red indicate little to no reduction (even worsening in some cases). This data must be viewed in conjunction with Table 4.1, as a significant percentage decrease is meaningless if the original spatial variance was already within an acceptable level.

Lastly, the musical samples were auditioned over Beyerdynamic DT770 Pro headphones [45] by one of the authors to provide an indication of audibility post-DiSP (Tables 4.3 and 4.4). This data must be observed in conjunction with the objective data since strong objective performance is worthless if the processing strongly colors the signal. Conversely, transparent processing that achieves little to no spatial variance reduction is also of no value.

4.1 Subwoofer only configuration

Considering the focus on practical applications of DiSP, the subwoofer only configuration data should be viewed as a purely academic exercise, as such a system (a subwoofer listened to in isolation) wouldn't

Spatial variance (dB)

		TDI duration		170 ms									341 ms									683 ms								
		Update rate		5 ms			10 ms			15 ms			5 ms			10 ms			15 ms			5 ms			10 ms			15 ms		
		Interp. factor	Unprocessed SV (dB)	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30
Subwoofer only (1 degree of freedom)	Cumulative analysis	50 ms	5.90	5.56	3.27	3.53	3.92	3.64	3.56	4.52	5.52	4.91	3.62	3.99	3.07	3.50	2.70	4.20	3.89	4.79	3.62	4.76	5.04	4.96	3.64	3.74	3.98	4.10	3.93	3.72
		270 ms	2.61	1.72	1.95	1.99	1.61	1.61	1.52	1.75	1.61	1.71	1.94	1.92	1.75	1.76	2.01	1.90	1.75	1.86	1.56	1.94	2.04	1.83	1.75	1.68	2.08	1.75	1.93	2.07
		5000 ms	2.01	1.93	1.67	1.69	2.20	2.29	1.76	1.62	1.47	2.18	1.83	2.04	1.83	1.95	1.26	1.38	1.73	1.79	1.42	1.71	1.61	1.47	1.86	1.92	2.10	2.35	1.67	1.66
	Partitioned analysis	Minimum	1.35	0.98	1.21	1.09	1.21	1.18	1.18	0.90	1.21	1.19	1.22	1.12	1.15	1.05	1.01	1.17	0.95	1.25	1.27	1.13	1.24	1.14	1.13	1.11	1.10	1.07	1.11	1.11
		Mean	2.05	1.74	1.74	1.71	1.68	1.69	1.70	1.72	1.75	1.69	1.71	1.72	1.69	1.67	1.69	1.75	1.67	1.72	1.73	1.74	1.73	1.72	1.74	1.76	1.76	1.70	1.71	1.69
		Maximum	3.19	2.79	2.25	2.33	2.53	2.47	2.29	2.71	2.63	2.37	2.56	2.40	2.26	2.50	2.50	2.54	2.61	2.51	2.47	2.59	2.57	2.45	2.63	2.69	2.68	2.48	2.52	2.46
Left, right + subwoofer (3 degrees of freedom)	Cumulative analysis	50 ms	6.82	4.78	3.88	4.53	3.74	3.81	3.09	4.75	7.42	5.44	4.78	4.22	5.64	3.52	4.79	4.25	4.04	5.00	3.86	5.91	3.50	4.18	4.94	3.97	3.75	4.10	4.38	6.60
		270 ms	6.52	1.82	2.52	1.91	1.92	1.54	1.66	1.85	2.80	1.98	2.25	1.70	2.27	2.20	1.97	2.16	2.26	1.89	1.86	6.90	2.01	1.97	2.24	2.22	1.90	2.34	1.94	2.08
		5000 ms	7.52	2.37	2.07	2.02	2.24	2.40	2.53	2.56	2.08	2.13	1.83	2.34	1.92	2.14	2.18	2.02	1.88	2.08	2.36	4.02	1.95	2.34	2.22	2.10	2.03	1.93	2.55	1.67
	Partitioned analysis	Minimum	1.71	1.34	1.23	1.28	1.35	1.45	1.20	1.44	1.40	1.44	1.25	1.31	1.36	1.39	1.40	1.17	1.32	1.50	1.25	1.41	1.19	1.37	1.45	1.34	1.32	1.21	1.43	1.30
		Mean	2.63	1.94	1.89	1.95	2.00	1.94	1.95	2.02	2.22	2.01	1.97	2.01	2.01	1.90	2.00	1.98	1.98	1.99	1.95	2.49	1.94	1.98	1.99	1.97	1.95	2.03	2.05	2.08
		Maximum	7.26	7.26	2.52	3.10	2.74	2.65	2.83	3.04	3.11	2.94	2.57	2.76	2.85	2.94	2.63	2.67	2.80	2.85	2.87	6.99	2.76	2.77	2.73	2.97	2.91	2.66	2.76	2.71

Table 4.1 Spatial variance measurements for the subwoofer only system and full system using cumulative analysis (at 50, 270 and 5000 ms) and partitioned analysis (with a 270 ms sliding window). All values are in dB.

Spatial variance (% change)

		TDI duration	170 ms									341 ms									683 ms								
		Update rate	5 ms			10 ms			15 ms			5 ms			10 ms			15 ms			5 ms			10 ms			15 ms		
		Interp. factor	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30	2	10	30
Subwoofer only (1 degree of freedom)	Cumulative analysis	50 ms	-5.8	-44.6	-40.2	-33.6	-38.3	-39.7	-23.4	-6.4	-16.8	-38.6	-32.4	-48.0	-40.7	-54.2	-28.8	-34.1	-18.8	-38.6	-19.3	-14.6	-15.9	-38.3	-36.6	-32.5	-30.5	-33.4	-36.9
		270 ms	-34.1	-25.3	-23.8	-38.3	-38.3	-41.8	-33.0	-38.3	-34.5	-25.7	-26.4	-33.0	-32.6	-23.0	-27.2	-33.0	-28.7	-40.2	-25.7	-21.8	-29.9	-33.0	-35.6	-20.3	-33.0	-26.1	-20.7
		5000 ms	-4.0	-16.9	-15.9	9.5	13.9	-12.4	-19.4	-26.9	8.5	-9.0	1.5	-9.0	-3.0	-37.3	-31.3	-13.9	-10.9	-29.4	-14.9	-19.9	-26.9	-7.5	-4.5	4.5	16.9	-16.9	-17.4
	Partitioned analysis	Minimum	-27.4	-10.4	-19.3	-10.4	-12.6	-12.6	-33.3	-10.4	-11.9	-9.6	-17.0	-14.8	-22.2	-25.2	-13.3	-29.6	-7.4	-5.9	-16.3	-8.1	-15.6	-16.3	-17.8	-18.5	-20.7	-17.8	-17.8
		Mean	-15.1	-15.1	-16.6	-18.0	-17.6	-17.1	-16.1	-14.6	-17.6	-16.6	-16.1	-17.6	-18.5	-17.6	-14.6	-18.5	-16.1	-15.6	-15.1	-15.6	-16.1	-15.1	-14.1	-14.1	-17.1	-16.6	-17.6
		Maximum	-12.5	-29.5	-27.0	-20.7	-22.6	-28.2	-15.0	-17.6	-25.7	-19.7	-24.8	-29.2	-21.6	-21.6	-20.4	-18.2	-21.3	-22.6	-18.8	-19.4	-23.2	-17.6	-15.7	-16.0	-22.3	-21.0	-22.9
Left, right + subwoofer (3 degrees of freedom)	Cumulative analysis	50 ms	-29.9	-43.1	-33.6	-45.2	-44.1	-54.7	-30.4	8.8	-20.2	-29.9	-38.1	-17.3	-48.4	-29.8	-37.7	-40.8	-26.7	-43.4	-13.3	-48.7	-38.7	-27.6	-41.8	-45.0	-39.9	-35.8	-3.2
		270 ms	-72.1	-61.3	-70.7	-70.6	-76.4	-74.5	-71.6	-57.1	-69.6	-65.5	-73.9	-65.2	-66.3	-69.8	-66.9	-65.3	-71.0	-71.5	5.8	-69.2	-69.8	-65.6	-66.0	-70.9	-64.1	-70.2	-68.1
		5000 ms	-68.5	-72.5	-73.1	-70.2	-68.1	-66.4	-66.0	-72.3	-71.7	-75.7	-68.9	-74.5	-71.5	-71.0	-73.1	-75.0	-72.3	-68.6	-46.5	-74.1	-68.9	-70.5	-72.1	-73.0	-74.3	-66.1	-77.8
	Partitioned analysis	Minimum	-21.6	-28.1	-25.1	-21.1	-15.2	-29.8	-15.8	-18.1	-15.8	-26.9	-23.4	-20.5	-18.7	-18.1	-31.6	-22.8	-12.3	-26.9	-17.5	-30.4	-19.9	-15.2	-21.6	-22.8	-29.2	-16.4	-24.0
		Mean	-26.2	-28.1	-25.9	-24.0	-26.2	-25.9	-23.2	-15.6	-23.6	-25.1	-23.6	-23.6	-27.8	-24.0	-24.7	-24.7	-24.3	-25.9	-5.3	-26.2	-24.7	-24.3	-25.1	-25.9	-22.8	-22.1	-20.9
		Maximum	0.0	-65.3	-57.3	-62.3	-63.5	-61.0	-58.1	-57.2	-59.5	-64.6	-62.0	-60.7	-59.5	-63.8	-63.2	-61.4	-60.7	-60.5	-3.7	-62.0	-61.8	-62.4	-59.1	-59.9	-63.4	-62.0	-62.7

Table 4.2 Percentage change in spatial variance for the subwoofer only system and full system using cumulative analysis (at 50, 270 and 5000 ms) and partitioned analysis (with a 270 ms sliding window).

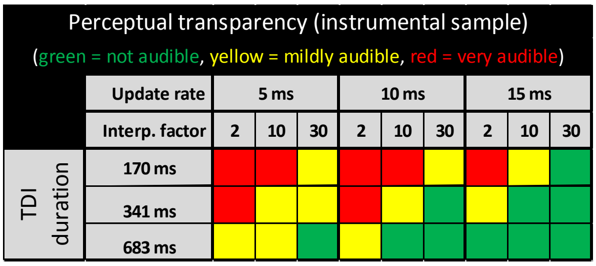

Table 4.3 Perceptual transparency of DiSP applied to the instrumental audio sample

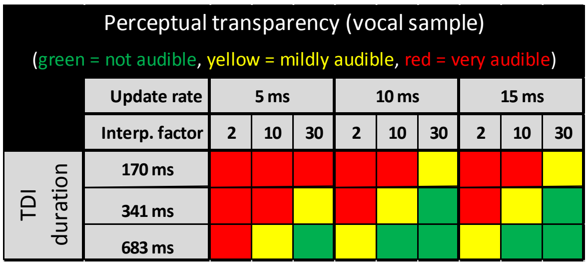

Table 4.4 Perceptual transparency of DiSP applied to the acapella vocal audio sample

be used in practice. Nonetheless, some interesting observations can be made.

Inspecting the unprocessed system's performance in Table 4.1 shows that the subwoofer naturally achieves low spatial variance (mean of 2.05 dB in the partitioned analysis). This is due to one of the authors specifically placing the subwoofer after an analysis of the room. The cumulative data shows, however, that short-term spatial variance reaches high levels (nearly 6 dB), meaning that transient-rich audio will sound different across the listening area. This problem is less severe over longer time analysis windows.

For the DiSP measurements, it appears that the shortest TDI duration (170 ms) doesn't provide consistent performance, where there are a number of situations where spatial variance worsens. This is likely due to the lack of frequency resolution in the TDI, resulting in poor decorrelation at the lowest analyzed frequencies.

Similarly, poor performance is observed in all instances with a 5 ms TDI update rate. This indicates that this is too rapid an update rate, which doesn't allow TDIs to be sufficiently isolated from adjacent TDIs. Additionally, a fast update rate requires significant processing power, so should be avoided.

In all cases with an interpolation factor of 2, performance is poor, indicating that a rapid shift from one TDI to the next has negative side effects, due to abrupt TDI transitions.

There are also poor results in most cases with an interpolation factor of 30, suggesting that such a gradual transition between TDIs defeats the purpose of dynamic DiSP, since there will be insufficient decorrelation between the direct sound and early reflections. This is in line with previous findings [33].

While there is very good performance with 683 ms TDIs, the performance isn't significantly better than 341 ms TDIs, so there is little justification for using longer TDIs; 341 ms appears to provide sufficient frequency resolution in the TDI generation process.

The informal subjective analysis presented in Tables 4.2 and 4.3, however, raises some caution of using 343 ms TDIs. There are few cases with the two chosen audio samples where the DiSP processing is completely transparent. There is slight coloration audible in many cases, where vocals appear to be more sensitive than instruments. If complete processing transparency is necessary, then the longer 683 ms TDI duration may be required. This is an area where further research is required.

Disregarding the potential (slight) audibility of the 341 ms TDIs, this leaves a scenario which could be ideal for such a DiSP implementation: TDI duration of 341 ms (at 48 kHz sampling rate), 10 ms TDI update rate (which is only slight longer than calculated as necessary for this particular configuration) and a TDI interpolation factor of 10. These settings will be the primary focus of the full system (left, right, subwoofer) analysis in the following section.

4.2 Left, right and subwoofer configuration

The full system configuration gives some interesting insights into the DiSP performance. First, the unprocessed system's cumulative analysis indicates that spatial variance is significantly worse than with the subwoofer only configuration. This is likely due to coherent interference between the three loudspeakers (they were all fed identical signals).

Interestingly, the partitioned analysis paints a much better picture than the cumulative analysis. It appears that the first 270 ms is especially poor in terms of spatial variance, as compared to later analysis frames.

It is an interesting juxtaposition in data, suggesting that after initial excitation spatial variance reduces, possibly due to the natural decorrelation of various reflections. Since typical audio content is dynamic in nature, though, the cumulative analysis may be more accurate as initial excitation by direct sounds and early reflections causes significant variance in listening experience. More research is required to reliably determine the significance of the differences seen between the two forms of data analysis.

What is clear with the full system is that the additional degrees of freedom provide better performance. While the subwoofer only DiSP gives mid- and long-term spatial variance reductions of 23.0% and 69.8%, respectively, the full system DiSP gives corresponding reductions of 29.8% and 71.0%, respectively.

The short-term performance is worse with the full system due to the lack of time-alignment between the stereo pair and the subwoofer. By the point of human hearing integration time, though, the full system has reduced spatial variance by nearly 70% as compared to 23% for the subwoofer only system.

This supports the suggestion that a 341 ms TDI with a 10 ms update rate and interpolation factor 10 is ideal. Furthermore, it is likely that full-range audio will partially mask DiSP artefacts (remembering that DiSP has only been applied below 250 Hz), resulting is an effectively transparent process.

5 Conclusions

It has been shown that with careful selection of DiSP variables a spatially-consistent low-frequency listening experience can be achieved when only using a standard single-subwoofer sound reproduction system in a domestic living room. Spatial variance can be further reduced (in many cases to imperceptible levels) with the inclusion of the stereo loudspeakers in the DiSP processing. In fact, it is evident that not doing so is likely to worsen performance significantly. This is something that few previous studies have investigated, but it appears to be of great importance for practical applications of such technology.

As with any form of audio processing, compromises are necessary. While improved transparency and performance can be achieved with longer TDIs, this requires significant processing power and adds what could be unacceptable amounts of latency to the audio stream. Similarly, it may seem germane to update TDIs as quickly as possible, but this requires greater processing power and can degrade performance due to significant overlapping of adjacent TDIs.

Based on the results from this study, guide DiSP settings for small room applications are as follows:

- TDI duration = 341 ms @ 48 kHz
- TDI update rate = 10 ms
- TDI interpolation factor = 10

Further research is necessary to investigate differences between the cumulative and partitioned analyses as well as into the application of partitioned convolution to minimize latency to allow for acceptable real-time implementation of DiSP (at the moment latency is unacceptably high).

Overall, this work shows that minimization of spatial variance in domestic listening scenarios is possible without any specialist hardware or calibration measurements. DiSP provides a turn-key solution to system optimization, which can be embedded entirely within a system's processor. It is a solution that would be accessible and acceptable to the general public.

References

[1] Toole, F.E. Sound *Reproduction: Loudspeakers and Rooms,* 1st edition. Focal Press, New York. 2008.

[2] Davis, M.F. "Loudspeaker Systems with Optimized Wide-Listening-Area Imaging". Journal of the Audio Engineering Society, vol. 35, no. 11, pp.888-896. November 1987.

[3] Olivieri, F., Fazi, F., Nelson, P., & Fontana, S. "Comparison of strategies for accurate reproduction of a target signal with compact arrays of loudspeakers for the generation of zones of private sound and silence." Journal of the Audio Engineering Society, vol. 64, no. 11, pp.905-917. December 2016.

[4] Shin, M., Fazi, F. M., Nelson, P. A., & Casagrande Hirono, F. "Controlled sound field with a dual layer loudspeaker array." Journal of Sound and Vibration, 333(16), 3794-3817. 2014.

[5] Hoffmann, F-M., & Fazi, F. "3D audio in rooms with compact wedge arrays." 4th International Conference on Spatial Audio, Graz, Austria, Graz, Austria. 2017.

[6] Menzies, D., & Fazi, F. "Room correction for object-based audio." Proc. Institute of Acoustics, Reproduced Sound. Southampton, UK. November 2016.

[7] Davis, D.; E. Patronis Jr. *Sound System Engineering:* Third Edition. Focal Press, Boston, MA. 2006.

[8] Backman, J. "Low-frequency polar pattern control for improved in-room response." 115th Convention of the Audio Engineering Society, paper 5867. October 2003.

[9] Welti, T.; A. Devantier. "Low-frequency optimization using multiple subwoofers." Journal of the Audio Engineering Society, vol. 54, no. 5, pp. 347-364. May 2006.

[10] Genereux, R.P. "Adaptive loudspeaker systems: correcting for the acoustics environment." 8th International Conference of the Audio Engineering Society. May 1990.

[11] Kuriyama, J.; Y. Furukawa. "Adaptive loudspeaker system." Journal of the Audio Engineering Society, vol. 37, no. 11, pp. 919-926. November 1989.

[12] 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.

[13] Bharitkar, S.; C. Kyriakakis. "Comparison between time delay based and non-uniform phase based equalization for multi-channel loudspeaker-room responses." 119th Convention of the Audio Engineering Society, paper 6607. October 2005.

[14] S. Bharitkar, C. Kyriakakis, "Objective Function and Automatic Multi-Position Equalization and Bass Management Filter Selection." 119th Convention of the Audio Engineering Society, paper 6608. October 2005.

[15] Santillan, A.O. "Spatially extended sound equalization in rectangular rooms." Journal of the Acoustical Society of America, vol. 110, no. 4, pp. 1989-1997. October 2001.

[16] Elliot, S.J.; L.P. Bhatia; D.W. Deghan; A.H. Fu; M.S. Stewart; D.W. Wilson. "Practical implementation of low-frequency equalization using adaptive digital filters." Journal of the Audio Engineering Society, vol. 42, no. 12, pp. 988-998. December 1994.

[17] Linkwitz, S. "Investigation of sound quality differences between monopolar and dipolar woofers in small rooms." 105th Convention of the Audio Engineering Society, paper 4786. September 1998.

[18] Vanderkooy, J. "Multi-source room equalization: Reducing room resonances." 123rd Convention of the Audio Engineering Society, paper 7262. October 2007.

[19] J. A. Pedersen, "Adjusting a Loudspeaker to its Acoustic Environment—The ABC System," 115th Convention of the Audio Engineering Society, paper 5880. October 2003.

[20] A. Celestinos, S. Birkedal Nielsen, "Low Frequency Sound Field Enhancement System for Rectangular Rooms Using Multiple Low Frequency Loudspeakers," 120th Convention of the Audio Engineering Society, paper 6688. May 2006.

[21] Hill, A.J.; M.O.J. Hawksford. "Practical applications of chameleon subwoofer arrays." 132nd Convention of the Audio Engineering Society, Budapest, Hungary. April 2012.

[22] Hill, A.J.; M.O.J Hawksford. "Low-frequency temporal accuracy of small-room sound reproduction." 133rd Convention of the Audio Engineering Society, San Francisco, CA, USA. October 2012.

[23] G. Kendall, "The Effects of Multi-Channel Signal Decorrelation in Audio Reproduction", ICMC Proceedings, pp. 319-326. 1994.

[24] Moore, J.B.; A.J. Hill. Applications of dynamic diffuse signal processing in sound reinforcement and reproduction. Proc. Institute of Acoustics – Conference on Reproduced Sound, Nottingham, UK. November 2017.

[25] D. Morgan, J. Hall, J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation", IEEE Transactions on Speech and Audio Processing, vol. 9, pp. 686-696. September 2001.

[26] M. Gerzon, "Signal Processing for Simulating Realistic Stereo Images," presented at the 93rd Convention of the Audio Engineering Society. October 1992.

[27] E. Stein and M. Walsh, "System and Method for Variable Decorrelation of Audio Signals." U.S. Patent No. 9,264,838. February 16, 2014.

[28] G. Potard and I. Burnett, "Decorrelation Techniques for the Rendering of Apparent Sound Source Width in 3D Audio Displays," Proceedings of the 7th Int. Conf. on Digital Audio Effects, pp. 280–284. October 2004.

[29] V. Pulkki, J. Merimaa, and T. Loki, "Reproduction of Reverberation with Spatial Impulse Response Rendering," 116th Convention of the Audio Engineering Society, paper 6057. May 2004.

[30] Moore, J.B. *Dynamic diffuse signal processing in sound reinforcement and reproduction.* Unpublished PhD Thesis, University of Derby. January 2019.

[31] M. O. J. Hawksford and N. Harris, "Diffuse Signal Processing and Acoustic Source Characterization for Applications in Synthetic Loudspeaker Arrays," 112th Convention of the Audio Engineering Society, paper 5612. April 2002.

[32] Moore, J.B.; A.J. Hill. "Dynamic diffuse signal processing for low-frequency spatial variance minimization across wide audience areas." 143rd Convention of the Audio Engineering Society, New York, USA. October 2017.

[33] Moore, J.B.; A.J. Hill. "Dynamic diffuse signal processing for sound reinforcement and reproduction." Journal of the Audio Engineering Society, vol. 66, no. 11. December 2018.

[34] Hill, A.J.; M.O.J. Hawksford. "A hybrid virtual bass system for optimized steady-state and transient performance." Proc. 2nd Computer Science and Electronic Engineering Conference (CEEC'10), IEEE Xplore, Colchester, UK. September 2010.

[35] Hill, A.J.; Hawksford, M.O.J.; Newell, P. "Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas: Effective and Practical Alternatives to Current Calibration Strategies." Journal of the Audio Engineering Society, vol. 64, no. 5, pp. 280-298. May 2016.

[36] Audio Physic, GmbH. Classic Compact loudspeaker. www.audiophysic.com

[37] NAD Electronics. D 3020 V2 Hybrid Digital DAC Amplifier. https://nadelectronics.com/product/d-3020v2-hybrid-digital-amplifier/

[38] Tannoy. TS2.8 Subwoofer. https://www.tannoy.com/Categories/Tannoy/TS2-SUBWOOFER/TS2-8-LO/p/P0C44#googtrans(en|en)

[39] Earthworks Audio. M30BX 30 kHz Battery Powered Measruement Microphone. https://earthworksaudio.com/products/microphones/measurement-series/m30bx/

[40] Sound Devices, LLC. USBPre. www.sounddevices.com

[41] Mathworks, Inc. MATLAB version 9.4 (R2018a). www.mathworks.com/products/matlab/

[42] Glasberg, B.R., Moore, B.C.J.; "Derivation of auditory filter shapes from notched-noise data." Hear. Res. 47, 103–138. 1990.

[43] Gelfand, S.A.; "Hearing: An introduction to psychological and physiological acoustics." CRC Press. 2017.

[44] Viemeister, N.F. and Wakefield, G.H.; "Temporal integration and multiple looks." Journal of the Acoustical Society of America, 90(2), pp.858-865. 1991.

[45] Beyerdynamic DT 770 Pro. https://europe.beyerdynamic.com/dt-770-pro.html