J. B. Moore and A. J. Hill, "Dynamic Diffuse Signal Processing for Sound Reinforcement and Reproduction,"
J. Audio Eng. Soc., vol. 66, no. 11, pp. 953–965, (2018 November.).
DOI: https://doi.org/10.17743/jaes.2018.0054

Dynamic Diffuse Signal Processing for Sound Reinforcement and Reproduction

JONATHAN B. MOORE, *AES Student Member,* **AND ADAM J. HILL,** *AES Member*
(j.b.moore@derby.ac.uk) (a.hill@derby.ac.uk))

Department of Electronics, Computing and Mathematics, University of Derby, Derby, DE22 1GB, UK

High inter-channel coherence between signals emitted from multiple loudspeakers can cause undesirable acoustic and psychoacoustic effects. Examples include position-dependent low-frequency magnitude response variation where comb-filtering leads to the attenuation of certain frequencies dependent on path length differences between multiple coherent sources, lack of apparent source width in multi-channel reproduction, and lack of externalization in headphone reproduction. This work examines a time-variant, real-time decorrelation algorithm for the reduction of coherence between sources as well as between direct sound and early reflections, with a focus on minimization of low-frequency magnitude response variation. The algorithm is applicable to a wide range of sound reinforcement and reproduction applications, including those requiring full-band decorrelation. Key variables that control the balance between decorrelation and processing artifacts such as transient smearing are described and evaluated using a MUSHRA test. Variable values that render the processing transparent while still providing decorrelation are discussed. Additionally, the benefit of transient preservation is investigated and is shown to increase transparency.

0 INTRODUCTION

In many sound reinforcement and reproduction scenarios, the desired audience sound coverage may only be achieved by the use of multiple electro-acoustic transducers emitting coherent signals at equal or nearly equal sound power levels. Where transducers are not arrayed in such a way that leads to acoustical coupling over their operational frequency range, any difference in path-length from a listening position or acoustic measurement point to two or more loudspeakers will result in a relative phase difference between the contributing signals [1]. The summed signal will have a frequency response that is dependent on the path-length differences, with cancellation of frequencies occurring where phase difference equates to 180 degrees.

For a two-transducer system, the fundamental frequency of cancellation is given by Eq. (0.1).

$$f_0 = \frac{1}{2t} \tag{0.1}$$

where, f_0 is the fundamental frequency of cancellation (Hz) and t is the time difference of arrival (TDOA) between the two transducers (s).

Therefore, f_0 is inversely proportional to TDOA, meaning that greater TDOA causes a lower fundamental frequency of cancellation. Additionally, any odd integer multiple of f_0 will also be subject to similar cancellation. This gives rise to the well-known comb-filtering effect.

When TDOA is small, comb-filtering is limited to mid- and high-frequencies. Subjectively, this is experienced as timbral anomalies between the received and source signals. For large-scale sound reinforcement systems, path-length differences are regularly on the order of several meters, leading to comb-filtering commencing at low-frequencies. In this case there will exist frequency-dependent amplitude nulls spanning several meters. The overall subjective implication is that audience members will receive a magnitude response that both differs from the source material and is position-dependent.

Spatial variance quantifies the magnitude response variation over a pre-determined frequency range and audience area [2–4] given by Eq. (0.2).

$$\mathrm{SV} = \frac{1}{N_f} \sum_{i\,=f_{lo}}^{f_{hi}} \sqrt{\frac{1}{N_p - 1} \sum_{p\,=\,1}^{N_p} \left(L_p(p,i) - \overline{L_p(i)}\right)^2} \tag{0.2}$$

where, SV is spatial variance in (dB), calculated based on the number of frequency bins (N_f), the number of measurement points (N_p), the frequency range of interest (f_{lo} to f_{hi}, in Hz), the sound pressure level (dB) at measurement point p and frequency bin i, $L_p(p,i)$, and the mean sound pressure

level (dB) over all measurement points at frequency bin i, $\overline{L_p(i)}$.

As described by Eq. (0.2), SV is the standard deviation of sound pressure level for all frequency bins of interest across all measurement points. 0 dB SV implies no deviation in magnitude response across a listening area, disregarding propagation loss if the responses are normalized to account for it.

A potential solution to high SV across an audience is to reduce inter-channel drive-signal coherence using decorrelation. In this case the acoustic signals will sum by the powers of their amplitudes since phase is randomized [5]. Should sufficient signal decorrelation be achieved, interference effects will be minimized and the resulting sound field will be approximately diffuse, characterized by a consistent magnitude response across an audience.

A signal decorrelation algorithm termed diffuse signal processing (DiSP), first described in [22], has been investigated in prior work by the authors [23–25]. It was found to be a useful tool for the decorrelation of multiple sources in sound reinforcement and reproduction applications. However, in [24] simulations showed that DiSP performance is reduced in closed acoustic spaces when decorrelation filters remain fixed. This is because direct sources maintain coherence with their early reflections, leading to comb filtering.

Therefore, a time-varying DiSP algorithm was introduced in [24], termed *dynamic DiSP*. This work advances dynamic DiSP with the introduction of two key user definable variables that may be used to balance dynamic decorrelation performance with processing perceptibility.

A MUSHRA style subjective test is presented to suggest suitable limits for these variables and to assess the transparency of the algorithm in comparison to unprocessed musical samples. Further to this, transient extraction prior to dynamic DiSP processing is utilized and its impact on decorrelation versus perceptual transparency is investigated.

After a brief review of existing signal decorrelation methods in Sec. 1, DiSP is reviewed in Sec. 2, followed by a justification for the need for a time-varying, dynamic variant of DiSP, capable of direct signal and early reflection decorrelation. Transient detection for the preservation of input signal's sharp transient content is also investigated (Sec. 3). The algorithm is objectively analyzed in Sec. 4 using image-source modeling. This is followed by subjective analysis of the algorithm's perceptual transparency in Sec. 5, where results of a multiple stimuli with hidden reference and anchor (MUSHRA) test are presented. A brief discussion on alternative applications of DiSP is given in Sec. 6 and the paper is concluded in Sec. 7.

1 SIGNAL DECORRELATION METHODS

Signal decorrelation algorithms have been described in previously published literature. Examples of applications for such algorithms are: generation of pseudo-stereo from monophonic sources [6–8], control of apparent source width [13], increased headphone externalization [29] and synthetic reverb [19]. Early algorithms were primarily used to produce a pseudo-stereophonic signal from a monophonic source. These methods rely on the generation of complimentary comb-filters by use of either delay lines or all-pass filters [6–8]. Using this method, two independent signals may be generated whose summed magnitude response is proportional to the magnitude response of the input signal. However, these methods are not suitable for the applications discussed in this work since only a limited number of sources may be decorrelated using this technique and perfect summation is only achieved in a limited sweet-spot [9, 10].

Other decorrelation methods have been developed for use in stereophonic echo-cancellation for voice conferencing [11, 12]. Unfortunately, these are also unsuitable for the specific sound reinforcement and reproduction applications in this research due to the level of distortion introduced and the limited number of decorrelated signals generated [13].

Kendall [14] describes a method of decorrelation filter generation, whereby filter coefficients are obtained via an inverse Fourier transform of a frequency domain specification of unity magnitude and random phase. This method allows for the generation of a large number of decorrelation filters that display low correlation with each other, allowing for many discrete sources to be decorrelated. However, it was found that while unity magnitude and random phase are specified at each frequency bin, the resulting magnitude spectrum from the inverse Fourier transform is not uniform in between these points leading to timbral coloration [14–16].

Decorrelation has also been achieved by passing a source signal through a filter bank to divide the it into critical frequency bands, with a random time shift applied to each band [15, 17, 18]. Depending on the magnitude of the random time shift calculated per frequency band, this method may result in frequency cancellation at band edges when the signal is reconstructed [15, 16]. This occurs when the time shifts equate to approximately 180 degrees phase difference between edge frequencies. This may be alleviated by constraining the time shift per band to multiples of 360 degrees phase shift for the edge frequency of each band [15]. However, this then limits the number of discrete sources that may be decorrelated. There will be a limited number of time shift values for each band that meet the criteria of being sufficient for decorrelation, not exceeding audible limits and still equating to a multiple of 360 degrees phase shift for the edge frequencies.

Spatial impulse response rendering (SIRR) is a method for multi-channel reproduction of measured room responses [19, 20]. A diffusion technique is necessary for the reproduction of diffuse sound over multiple loudspeakers. Initially, diffusion is achieved by creating continuous uncorrelated noise for each loudspeaker. Using a short-term Fourier transform, the magnitude of each time–frequency component of each noise signal is set equal to the magnitude response of the source signal. In an investigation that applied the technique to directional audio coding (DirAC), the method was found to be inadequate due to the distortion and pre-echoes produced [21]. Instead, a method of

convolution with exponentially decaying white noise bursts was used. This method is similar to the one proposed in [22]. In both works it is noted that to achieve adequate low-frequency decorrelation, long noise bursts must be used, however at high-frequencies this causes perceptual issues such as transient smearing.

Diffuse Signal Processing (DiSP) [22] describes the synthesis of impulses with rapidly decaying random phase noise tails, termed temporally diffuse impulses (TDIs). To achieve system decorrelation, each discrete source drive-signal is convolved with a unique TDI.

In TDI synthesis, an exponential decay along with a random phase shift is applied to each frequency component. Applying a longer exponential decay to a given frequency component during TDI synthesis results in greater reduction of inter-channel coherence at that frequency bin, at the expense of increased filter audibility. Manipulation of exponential decay constants by frequency component allows for enhanced control over the level of decorrelation achieved versus perceptual impact across the spectrum. This method is particularly attractive as exponential decay constants can be optimized to provide sufficient low frequency decorrelation while minimizing audible effects such as transient smearing at higher frequencies. Additionally, the technique is easily scalable to an arbitrary number of discrete sources as all that is required is a unique TDI for each source in the system. Therefore, DiSP forms the basis of the algorithm described in this work.

2 STATIC DIFFUSE SIGNAL PROCESSING

The synthesis of TDIs was first described in [22], where TDIs remain fixed over the entirety of the system's operation, hence in this work the approach is referred to as *static DiSP*. This section summarizes TDI synthesis, described originally in [22], with optimization techniques novel to this work.

2.1 TDI Generation

Each TDI is synthesized from the summation of exponentially decaying, random phase cosine waves of increasing frequency up to the Nyquist frequency, defined as the highest frequency that may be sampled without causing aliasing at the specified system sample rate.

TDI length may be defined by the user, however informal testing has shown that for an audio sample rate of 44.1 kHz a length of at least 8192 samples is necessary to provide adequate frequency domain resolution to decorrelate down to 20 Hz. The TDIs used in this work are of 8192 samples length for use at 44.1 kHz sample rate.

The decay rate of each cosine wave is determined by a pre-defined time constant in seconds, which is then converted to decay constant by Eq. (2.1).

$$\mathrm{DC}(n) = \sum_{n=0}^{\frac{N}{2}-1} \frac{N}{\mathrm{TC}(n) \times F_s} \tag{2.1}$$

where, N is the TDI length in samples, DC and TC are vectors of length *N/2* containing the decay and time constants, respectively, and F_s is the sample rate (Hz). The variable n represents the frequency bin index under inspection.

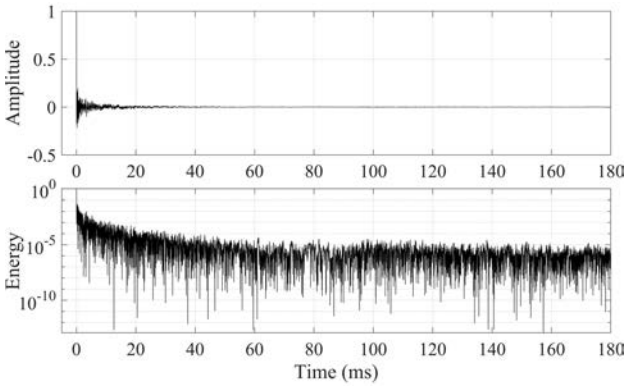

Fig. 1. Time domain representation of an example TDI of length 8192 samples at 44.1 kHz sample rate showing the initial impulse followed by rapidly decaying noise tail. Amplitude over time (above), and energy over time (below).

The phase of each frequency component is obtained using a random sequence of values between $\pm\pi$ with 0.94 weighting, which limits the randomized phase values to $\pm 0.94\pi$, as suggested in [22]. This weighting has been found to be important with regard to ensuring the initial impulse of the TDI occurs at time zero. The random phase values are generated and stored according to Eq. (2.2).

$$P(n) = \sum_{n=0}^{\frac{N}{2}-1} 2\pi\,(R(n) - 0.5)\,W \tag{2.2}$$

where, P is a vector of length *N/2* containing all processed phase values, R is a vector of uniformly distributed random numbers between 0 and 1 of length *N/2*, and W is the phase weighting factor, equaling 0.94 in this case. TDIs are then synthesized using Eq. (2.3).

$$\mathrm{TDI} = \frac{1}{N} \sum_{n=0}^{\frac{N}{2}-1} \frac{\cos\left(\frac{P(n)+2\pi rn}{N}\right) e^{\left(\frac{-\mathrm{DC}(n)r}{N}\right)}}{\sigma\left(\cos\left(\frac{P(n)+2\pi rn}{N}\right) e^{\left(\frac{-\mathrm{DC}(n)r}{N}\right)}\right)} \tag{2.3}$$

where the temporally diffuse impulse (TDI) is generated based on a summation of cosine waves at frequencies from zero (DC) to Nyquist frequency and r is a vector of length N with values spaced linearly from *0* to *N – 1*. In Eq. (2.3) σ represents the standard deviation operator. Each cosine wave added to the composite TDI is normalized to its standard deviation so that each frequency component carries equal energy. Without this, phase randomization could result in an inconsistent summation across the frequency band [22].

An all-pass response for each TDI is achieved via minimum phase equalization, as described in [22]. Fig. 1 shows the time domain representation of an example TDI.

Each TDI generated exhibits a different phase response due to the random phase generation process. All other variables, such as TDI length and time constant for each frequency component remain fixed. Therefore, when multiple TDIs interact, overall system performance can be defined

by manipulation of the frequency-dependent time constants prior to TDI generation. These control the decay time for each individual frequency component in TDI synthesis. Longer frequency decay times lead to greater reductions in inter-channel coherence at the expense of increased filter transient smearing, while shorter frequency decay times lead to reduced decorrelation performance with increased processing transparency.

Previously published work determined that a uniform probability density function (PDF) was ideal for use in random phase generation with time constants following a linear relationship inversely proportional to frequency [22]. Recent research by the authors [23, 24] established that uniform PDF performance can be improved with a nonlinear time constant relationship where time constants are manually defined by octave band. This gives optimal performance with regard to the decorrelation achieved with minimal perceptual degradation. The optimization of TDI generation for achieving maximal decorrelation while minimizing perceptual effects is discussed in the following section.

2.2 TDI Optimization

Previously published work suggests that time constants should be defined for the highest and lowest frequencies with intermediate values interpolated via a linear or logarithmic function [22]. This is based on the assumption that decay times should be inversely proportional to frequency. However, informal subjective assessments by the authors revealed that when only defining the highest and lowest frequency decay times, it is difficult to achieve sufficient low-frequency decorrelation without introducing noticeable temporal effects at mid- and high-frequencies. This is especially noticeable with transient-rich material. Enhanced control over time constant versus frequency is required.

It is suggested in [23, 24] that defining decay time constant by octave band allows for selection of a TDI frequency dependent decay characteristic that is more in line with human perception. In this work the "audible threshold of decay time constant" is defined as the time constant value at which a TDI becomes audible for a given band when all other frequencies are passed without effect. A subjective test was developed by the authors to obtain the audible threshold of decay time constant for the frequency bands defined in Table 1 using transient source material, which has been found to be most revealing—in this case drum loops were used [24]. The results are summarized in Table 1.

To integrate this data into TDI generation, the central frequency of each band is set to the decay time constant given in Table 1. Intermediate time constants are obtained via linear interpolation between the central band points [24]. The change of decay time constant over frequency obtained with this method is shown in Fig. 2 with comparison to the linear and logarithmic methods described in [22]. When comparing the variable decay method curve, obtained by the audible decay time thresholds in Table 1, to the linear and logarithmic curves, it becomes clear that mid- to high-frequency time constants derived from the linear and logarithmic methods exceed audible limits. Therefore, the values obtained for the variable decay constant method in [24] will be used in this work.

Table 1. Results of subjective test for audible threshold of decay time constant by frequency band [24].

Frequency band (Hz)	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

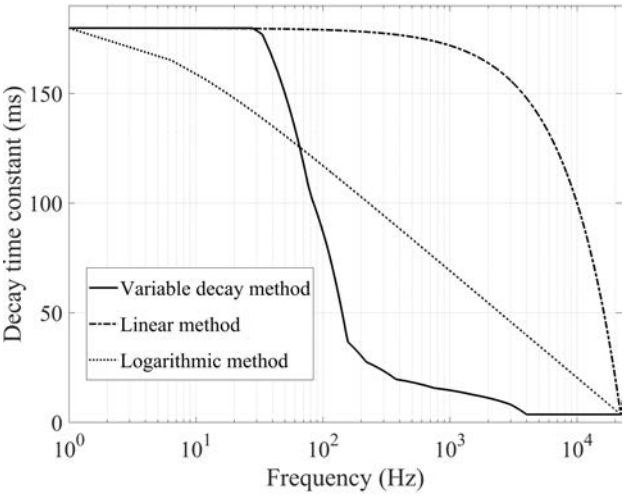

Fig. 2. Comparison of the decay time vs. frequency relationship obtained by the logarithmic and linear methods [22] and the variable decay method using subjectively-obtained audible limits [24]

Another aspect of TDI optimization for consideration, which is closely linked to time constant selection, is that of the amplitude of the noise tail in comparison to the initial impulse of the TDI. While TDIs will decorrelate up to the Nyquist frequency, in real-world applications this is unlikely to be necessary. However, decay times for frequencies above which decorrelation is desired still need to be considered as their selection impacts the amplitude of the noise tail in comparison to the initial impulse, and therefore the level of decorrelation achieved over all frequencies. If the amplitude of the noise tail is greatly reduced, very little decorrelation will be achieved. If the amplitude of the noise tail is increased, greater decorrelation will be achieved at the expense of increased filter audibility. When audible filters are used, the audio will sound as if a short decay reverb has been applied and transients will be smeared.

Informal subjective evaluations show that the choice of 3.7 ms time constant for all frequencies above 4 kHz can cause audible artifacts such as resonances or ringing for highly-transient source material. This is resolved by setting

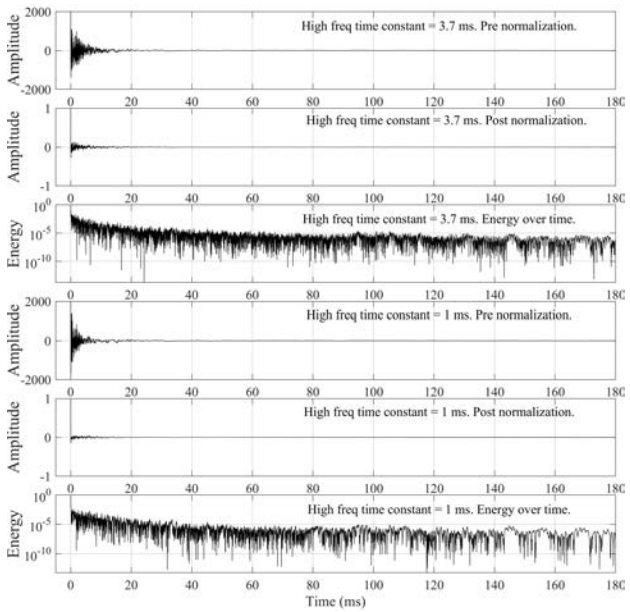

Fig. 3. Difference in initial impulse/noise tail amplitude ratio, pre- and post-normalization with identical TDI parameters apart from time constant above 4 kHz (high frequency time constant), along with the resultant difference in TDI energy over time when the high frequency time constant is reduced from 3.7 to 1 ms.

the time constant for all these frequencies to an arbitrarily short time constant (1 ms)—as a result, minimal decorrelation is achieved at these high frequencies but the unwanted artifacts are eliminated. A side effect is that the initial impulse is increased in amplitude in relation to the noise tail, as the energy from these high-frequency components only significantly contributes to the initial impulse, not the noise tail. This means that less decorrelation is achieved for the full spectrum, despite all other time constants remaining fixed. The effect that reducing the high frequency time constant has on noise tail amplitude is illustrated in Fig. 3. All other decay time constants are as in Table 1.

This aspect of TDI generation can be controlled by a single variable, termed high frequency time constant. The level of decorrelation versus perceptual impact of the processing may be adjusted while other parameters can remain fixed. Therefore, in this work TDIs generated using the time constants obtained in [24] must be investigated with different high-frequency time constants (>4kHz). High-frequency time constants of 1 ms, 3.7 ms, and 6.4 ms are investigated objectively in Sec. 4 and subjectively Sec. 5.

2.3 Performance Limitations

Prior work by the authors has shown that static DiSP, where the decorrelation filters remain fixed over time, gives reduced performance when applied in closed acoustic spaces [23, 24]. This is due to the interaction of surface reflections of the same source origin producing comb-filtering. In this work *dynamic DiSP* is investigated, where TDIs are changed over the course of milliseconds to decorrelate system sources from their own output over time. In the dynamic system sources are not only decorrelated from each other but also their own early reflections. Provided the rate of change of TDIs is sufficient for a given acoustic topology, there should be a reduction in magnitude response variation in enclosed acoustic spaces that static DiSP is unable to achieve.

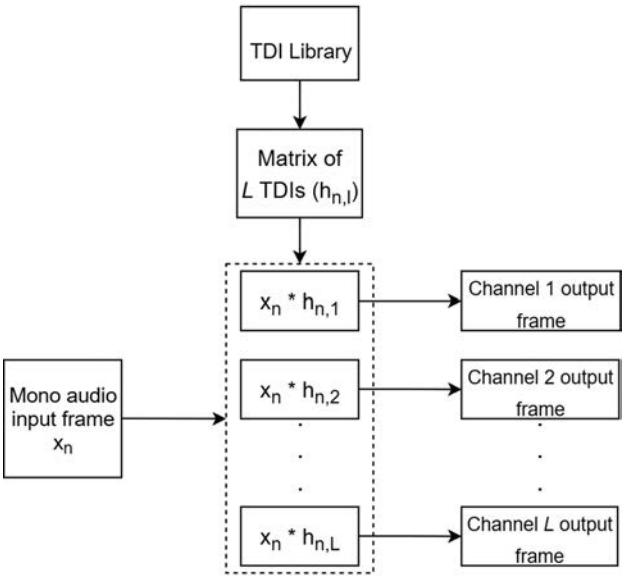

Fig. 4. Diagram of the dynamic DiSP algorithm

This was verified in a series of real-world measurements in prior work by the authors [25]. Low-frequency spatial variance reduction was measured for two systems—a small domestic room and a medium-sized music venue. Dynamic DiSP was shown to outperform static DiSP in all cases. There are, however, perceptual concerns associated with rapidly changing a source's TDI. The next section describes the dynamic DiSP algorithm and how to mitigate any perceptual issues.

3 DYNAMIC DIFFUSE SIGNAL PROCESSING

The TDI generation algorithm described in Sec. 2 allows for the creation of an arbitrary number of decorrelation filters. As all parameters apart from phase generation remain fixed, filter audibility and system performance are predictable for a given set of input parameters. For a system comprising of L discrete sources, for each mono audio input frame, L TDIs are drawn from a pre-generated library. Each TDI is convolved with the audio frame generating L decorrelated channels, which are outputted to the L discrete sources. This process is illustrated in Fig. 4.

Unless an overlapping sliding output window is used, changing TDI coefficients from one output frame to the next results in audible clicking. Therefore, a sliding overlapping output window of 1/3rd the output frame length is used. In this way, each output frame per source is processed by three overlapping distinct TDIs and the audible clicking is eliminated. Additionally, interpolation of TDI coefficients is used to smooth the TDI transition. This is detailed in Sec. 3.2.

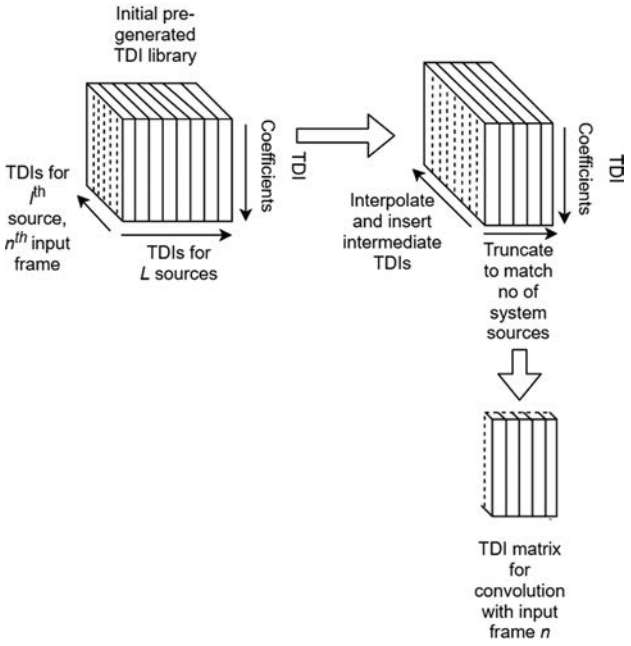

Fig. 5. Diagram showing how the initial pre-generated TDI matrix is handled given user inputs of number of transducers and interpolation factor

The rate at which TDIs must be updated is defined by the acoustic and system topology. For large spaces the rate may be relaxed as the path length difference between direct source and early reflections increases. For small spaces (e.g., domestic rooms) the maximum effective TDI update rate can often be less than 10 ms.

3.1 TDI Update Rate Calculation

It is key that the TDI update rate is fast enough so that a source's direct sound and first arriving reflection at a listening location are each processed by a different TDI. The necessary TDI update rate is dependent on the room size and system configuration. For practical purposes, the maximum rate is calculated using a simplified geometrical calculation [26]. The shortest reflection path length that corresponds to half a wavelength delay of the highest frequency of interest at a central measurement location is found from the room dimensions and source position. The required TDI update rate is then found with Eq. (3.1).

$$dT = 1000\frac{\Delta l}{c} \tag{3.1}$$

where, dT is the required TDI update rate (ms) and is calculated based on the path length difference (m) between the direct sound and first-arriving problematic reflection, Δl and the speed of sound, c (m/s).

3.2 TDI Library Configuration and Interpolation

In this work TDI libraries are pre-generated to handle up to 20 discrete transducers. For each transducer, 100 initial TDIs are generated and stored in a matrix as shown in Fig. 5. With dynamic DiSP, once the final set of TDIs has been drawn from the library, the first TDI set is drawn again and the process repeats.

Informal testing has shown that when fast TDI update rates are necessary (<10 ms) the changing of filter coefficients becomes audible (perceived as a "phasing" sound) despite overlapping output windows. This can be alleviated with the generation of intermediate TDIs via linear interpolation of TDI coefficients. It is key that the minimum phase equalization stage of TDI generation occurs *after* any interpolation in order to ensure an all-pass response for all TDIs generated.

The effect of generating intermediate TDIs is a reduction in audible effects of changing filters as the transitions are smoothed, but there is a negative impact on dynamic decorrelation performance due to the increased similarity between consecutive filters. However, the discrete channels are still decorrelated from each other as with static DiSP.

Dynamic DiSP performance may be controlled by an interpolation factor that defines the number of interpolation points between pre-generated TDI coefficients. The user may input the desired interpolation factor and the TDI library is adjusted accordingly before beginning real-time processing, as shown in Fig. 5, thus giving no impact to real-time processing efficiency.

3.3 Transient Handling

For adequate low-frequency decorrelation, long decay times are necessary. It has been found that a TDI length of 8192 samples at 44.1 kHz is required to give sufficient low-frequency decorrelation down to 20 Hz. This equates to a filter duration of 185.7 ms. The audible effect of using such long filters is mitigated primarily by the frequency-dependent exponential decay. As shown in Table 1, most of the frequency components persist for a much shorter duration. However, the necessarily long exponential decays may lead to temporal smearing depending on the amplitude of the TDI noise tail. This is illustrated in Fig. 6 which shows the effect of the dynamic DiSP algorithm on short transients with increasing high frequency decay constant. Increasing the high frequency decay constant increases the amplitude of the TDI noise tail.

The use of transient detection in decorrelation algorithms has been discussed in [7, 18, 27]. It is thought that by constraining decorrelation to only what is deemed the steady-state portion of the signal, temporal smearing may be alleviated without significant impact to low-frequency decorrelation due to the short duration of the extracted transients.

In this work the transient extraction method utilizes a constant-Q transform (CQT) due to its superior low-frequency resolution to the discrete Fourier transform [28]. Optimization of the transient detection stage for real-time processing is the subject of future work.

The mono input signal is transformed into the frequency domain via the CQT, given by Eqs. (3.2) and (3.3) [27].

$$N_k = \frac{F_s Q}{F_k} \tag{3.2}$$

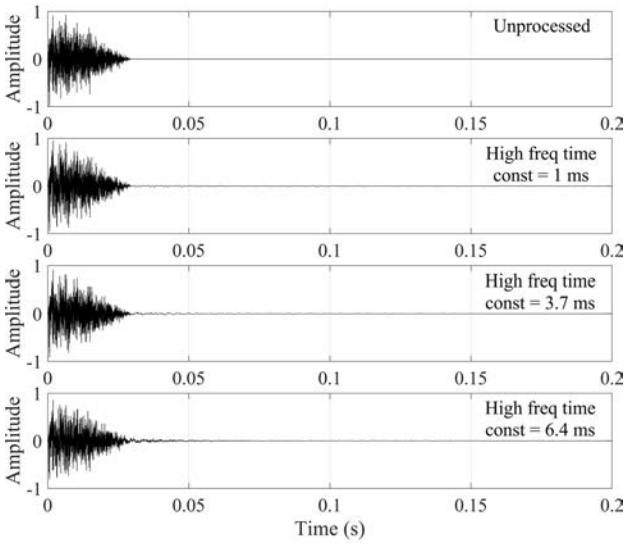

Fig. 6. The effect of the dynamic DiSP algorithm on short transients when processed using a TDI length of 185.7 ms. The time constants used are as in Table 1 apart from the high frequency time constant as shown in the top right of each plot.

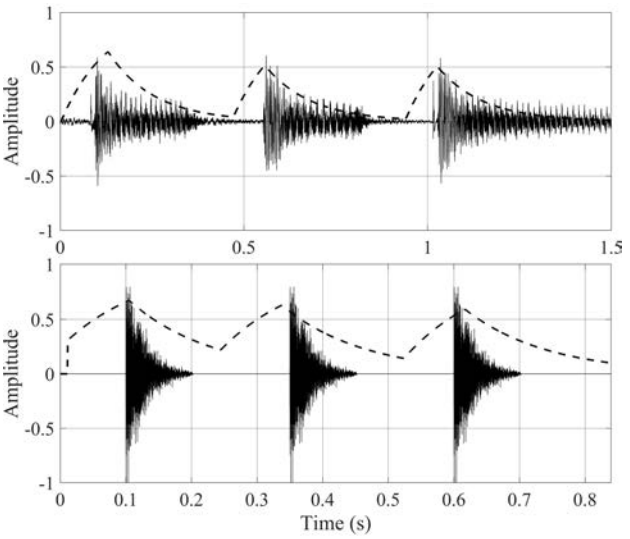

Fig. 7. Example transient detection weighting function (dashed curve) superimposed over drum loop (above) and short transient inputs of 0.1 s (below). Weighting moves towards 1 when a transient is detected

$$X_k = \frac{1}{N_k} \sum_{n=0}^{N_k-1} w_k(n)\, x(n)\, e^{\frac{-j2\pi Qn}{N_k}} \tag{3.3}$$

where, N_k is the required analysis window length in samples at frequency bin k, F_s is the sample rate (Hz), f_k is the frequency at the k^{th} bin, X_k is the CQT of the signal, w_k is the windowing function of the input signal, in this case a Hann window, x is the input signal, and Q is the required ratio of frequency to spectral resolution. The frequency dependent term, N_k, allows for an adaptive analysis window size, giving a constant resolution to center frequency ratio.

The transient detection algorithm outputs a weighting function with values between 0 and 1, changing over time. This output is generated by monitoring spectral energy content using the CQT. If the change in spectral energy between successive windows exceeds a pre-defined threshold, the weighting function moves towards 1 indicating transient material, or moves gradually towards zero when this threshold is not exceeded (Fig. 7).

The input signal is then transformed into a transient signal by multiplication of the weighting function with the input signal. The steady-state signal is subsequently obtained by subtracting the transient signal from the input signal. The steady-state signal is passed through the dynamic DiSP algorithm, then summed with a delayed copy of the transient signal to give the final output.

4 OBJECTIVE EVALUATION

In the dynamic DiSP algorithm, the key parameters for controlling the balance between decorrelation and perceptual effects are: the choice of interpolation factor, which smooths TDI transition; the choice of high-frequency decay time constant, which dictates the ratio of diffuse noise tail to initial impulse amplitude in TDI generation; and the application of a transient extraction algorithm. The first objective evaluation, therefore, must examine the interaction between these three parameters to judge how they affect dynamic DiSP performance.

4.1 Testing Method

For this analysis TDI libraries were pre-generated, as described in Secs. 2 and 3. TDIs were of length 8192 samples, with an audio sample rate of 44.1 kHz giving a TDI duration 185.7 ms. The time constants used matched those shown in Table 1, apart from high-frequency time constant (>4000 Hz) choices of 1 ms, 3.7 ms, 6.4 ms, and 9.1 ms. Interpolation factor was also varied with choices of 0, 10, 20, and 30 indicating how many intermediate TDIs were to be interpolated between each successive pair of pre-generated TDIs. Ten TDI libraries covering each combination of conditions were generated. The results presented are the average performance over the 10 libraries generated for each combination of conditions.

While dynamic DiSP is applicable to any frequency range, the algorithm's performance was investigated here with regard to large scale low-frequency live sound reinforcement. In this example, decorrelation is applied to a 4-source subwoofer array for the reduction of low-frequency spatial variance between 20–250 Hz. An image source model was used as in [23, 24] to simulate a 24 m × 30 m × 18 m space, which reflects a typical medium/large-scale venue.

All surface absorption coefficients were set to 0.2. Reflections up to 15th order were modeled with four point-sources positioned at (x, y) coordinates (2.4 m, 3 m), (5.6 m, 3 m), (18.4 m, 3 m), and (21.6 m, 3 m) all located 1 m off the ground. A 100-point measurement grid was positioned centrally within the space with a point-to-point spacing of 1.6 m. A musical signal, *Tom Sawyer* by Rush, was used to

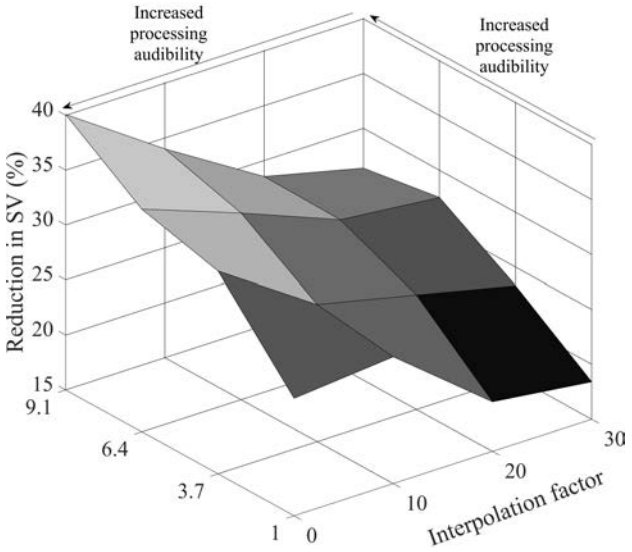

Fig. 8. Reduction in SV for all test conditions *without transient detection*

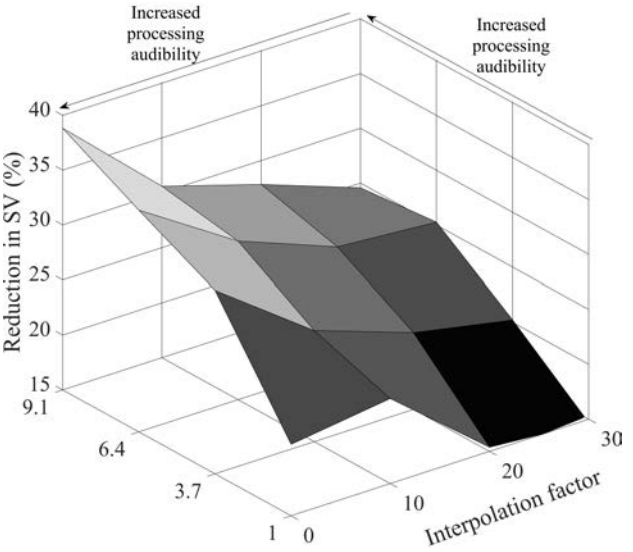

Fig. 9. Reduction in SV for all test conditions *with transient detection*

excite the space with the audio signal processed at a TDI update rate of 25 ms.

The complex frequency response of the summed signal at each measurement point was taken using a fast Fourier transform (FFT). The transfer function of each measurement point was obtained by dividing the measured response by the FFT of the delayed input signal. The magnitude response of each measurement point was extracted and smoothed by $1/10^{th}$ per octave to closer match human perception than typical $1/3^{rd}$ octave smoothing [30]. The 20–250 Hz FFT bins of each measurement point response were then used in Eq. (0.2) to calculate spatial variance (SV) over a 1.0 s analysis window. The initial, unprocessed SV of the modeled system was 3.4 dB.

4.2 Results

The results of the processing for all conditions, with and without transient detection, are given in Figs. 8–10. As expected, dynamic DiSP performance is reduced by increasing the interpolation factor, which increases the level of TDI transition smoothing, and decreasing the high-frequency decay time constant, which reduces the peak amplitude of the random phase noise tail. The addition of transient detection does not reduce performance significantly. A 2.4% mean decrease in performance by the addition of transient detection was seen over all test cases. While Figs. 8 and 9 show that a significant reduction of low-frequency spatial variance is possible with this processing, audible limits for the processing must be investigated. A MUSHRA [28] subjective test was performed to assess this and is described in Sec. 5.

To further illustrate the effects of dynamic DiSP, the smoothed low-frequency magnitude responses for the unprocessed and dynamic DiSP processed systems are shown in Fig. 11. The dynamic DiSP processed system had a high-frequency decay time constant of 3.7 and an interpolation factor of 20, giving an SV reduction of 25%.

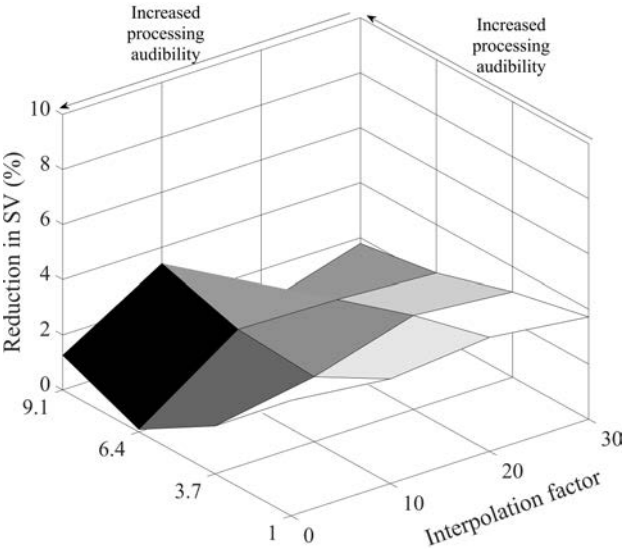

Fig. 10. Difference in performance between transient and non-transient detected dynamic DiSP

5 SUBJECTIVE EVALUATION

The aim of the subjective evaluation was to assess the perceptual transparency of the dynamic DiSP algorithm using parameters that would be applicable to a variety of sound reinforcement and reproduction applications. The test was performed in a hemi-anechoic chamber built in accordance with ISO 26101, with subjects undertaking the test twice, once over a pair of open-back Beyerdynamic DT 770 headphones and once over a d&b audiotecknik Y7P and a d&b B Subwoofer system, both on-axis to the listener.

The reason for choosing this test set up and not replicating the setup used for the objective test is that in such real-world scenarios the true processing transparency of the algorithm may not be accurately assessed due to masking

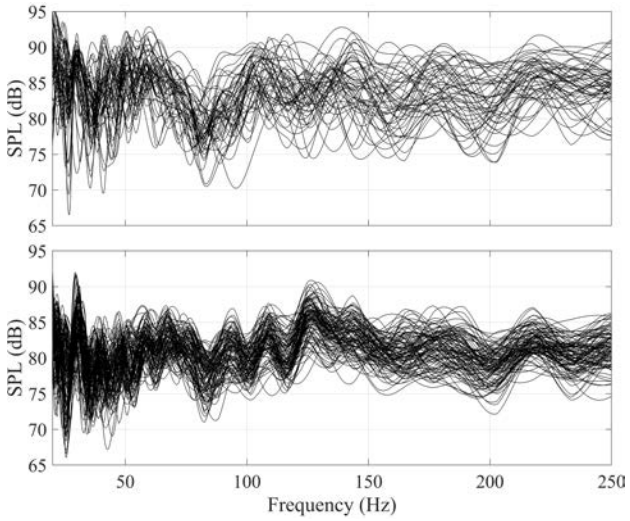

Fig 11. 1/10th octave smoothed magnitude responses of 100 measurement positions across a 24 m × 30 m audience area excited by 4 point-sources for the unprocessed system (top) and the dynamic DiSP-processed system (bottom)

effects of room acoustics. Additionally, it is also important that the algorithm be assessed when only a single mono source is presented. This is because in large scale sound reinforcement, it is not uncommon for some audience positions to be predominantly covered by a single source, or a cluster of coupled sources, such as at the edges of an audience area. In such a case the decorrelation algorithm should not rely on the contributions of other discrete sources for transparency.

Twenty-eight subjects participated in the MUSHRA test to subjectively evaluate the perceptual impact of dynamic DiSP. Seventeen of the subjects had prior listening test experience. The participants were between the ages of 20 and 37, consisting of 25 males and 3 females. All participants reported having healthy hearing. For each subject the order of presentation method (headphones or loudspeakers) was alternated.

As per the guidelines in [28], each subject assessed the subjective audio quality of eight 10 s audio samples that were identical apart from the type of dynamic DiSP processing applied, in comparison to an unprocessed reference signal.

The test was repeated with three musical signals which were presented in random order: a rock piece, a dance piece, and a rap piece. The pieces were respectively: *Tom Sawyer* by Rush, *Disco Drive* by DJ Qness, and *Lovin' It* by Camp Lo. The test samples contained a hidden reference identical to the original audio signal, a low anchor, and six processed signals under test. The TDI update rate for the test material was 15 ms. It has been found that this rate is sufficient for room sizes down to around 5 m × 4 m × 3 m when seeking to correct low-frequency spatial variance up to 250 Hz. Faster TDI update rates may introduce further distortion which is not evaluated here.

The parameters for dynamic DiSP processing to be tested were informed by the objective test in Sec. 4. The objective tests show a clear trend that increasing high frequency decay constant, which increases noise tail peak amplitude, and decreasing interpolation factor, which reduces the level of dynamic TDI transition smoothing results in increased SV reduction (Figs. 8–9). It is expected that this is at the cost of increased filter audibility and reduction in audio quality. Therefore, the independent variable for the subjective test was chosen to be the level of dynamic DiSP performance as dictated by these two variables. Table 2 shows the values selected for each of the test samples, and the SV reduction performance of these values can be seen in Figs. 8 and 9.

Table 2. Description of MUSHRA test samples.

Test sample	Processing parameters (High frequency time constant, interpolation factor)
1 (Hidden ref)	Unprocessed
2 (Low anchor)	9.1 ms, 0 (w/o transient extraction)
3	1 ms, 30 (w/o transient extraction)
4	1 ms, 30 (w/transient extraction)
5	3.7 ms, 20 (w/o transient extraction)
6	3.7 ms, 20 (w/transient extraction)
7	6.4 ms, 10 (w/o transient extraction)
8	6.4 ms, 10 (w/transient extraction)

The TDIs used were of the same length as in the objective test (8192 samples with an audio sample rate of 44.1 kHz) and utilized the decay time constants shown in Table 1, apart from the high frequency time constant, which was varied as shown in Table 2. The TDIs provided full-spectrum processing, however, since low-frequency decorrelation is of specific interest here, the test audio was only processed up to 4 kHz using a complimentary low-pass/high-pass stage with crossover set at 4 kHz. Only the low-passed signal was processed. Without this stage there may be further perceptual effects that are not assessed here, but the results shown are applicable for dynamic DiSP of TDI update rates $\geq$ 15 ms and decorrelation from 20–4000 Hz, which are sufficient parameters for most applications.

For each test the presented sample order was randomized. The test signals are described in Table 2.

The low anchor used was *not* the bandlimited anchor described by [28]. It has been found in informal listening that the dynamic DiSP parameters shown for test sample 2 in Table 2 produced a large amount of distortion, and it was decided the sample would be more suitable for a low anchor in this test. All subjects correctly identified the low anchor.

The GUI presented to the subjects is shown in Fig. 12. The final scores of each subject were normalized in the range 0 to 100, where 0 corresponds to the bottom of the scale, or "bad sound quality" as described in [28].

5.1 Results and Analysis

Fig. 13 shows the overall MUSHRA scores obtained for each source material.

As expected, test clips with greater levels of dynamic DiSP processing as defined by Table 2 scored lower in terms of audio quality. Additionally, the samples where transient detection was incorporated into the

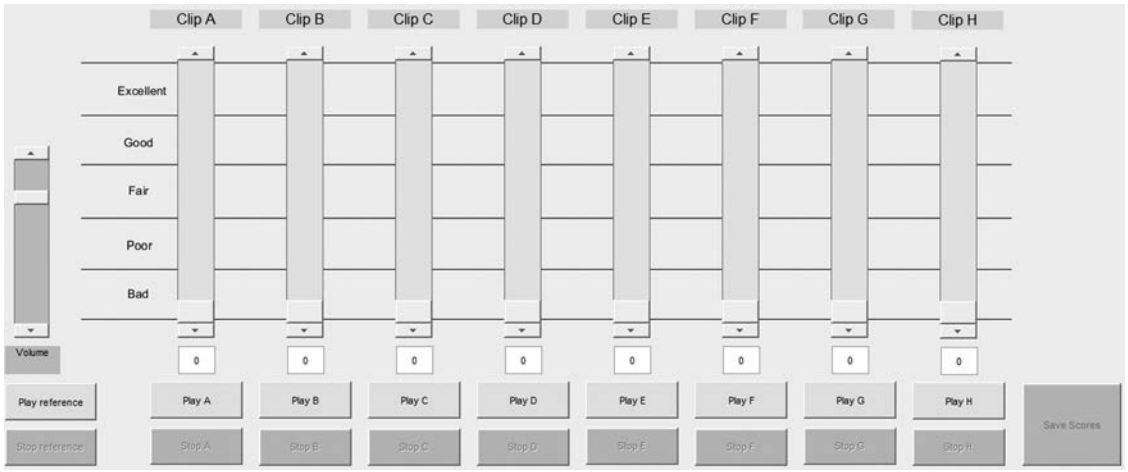

Fig. 12. MUSHRA GUI presented during the subjective evaluation of dynamic DiSP

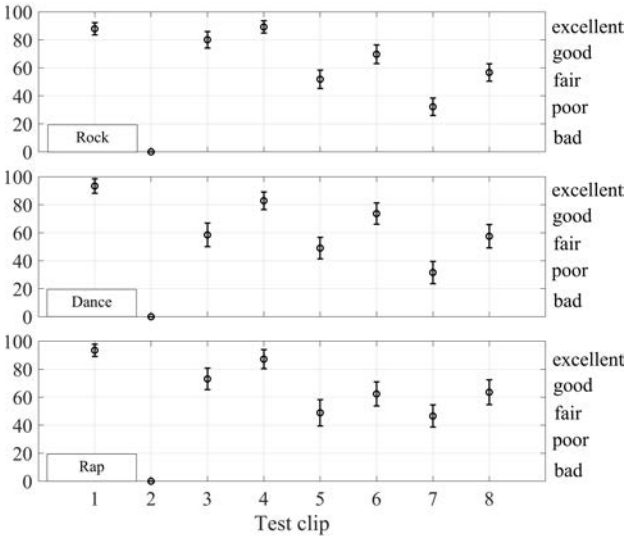

Fig. 13. Overall MUSHRA test results for different audio samples. Mean scores shown with 95% confidence intervals for six different test materials and high/low anchors, numbered as in Table 2

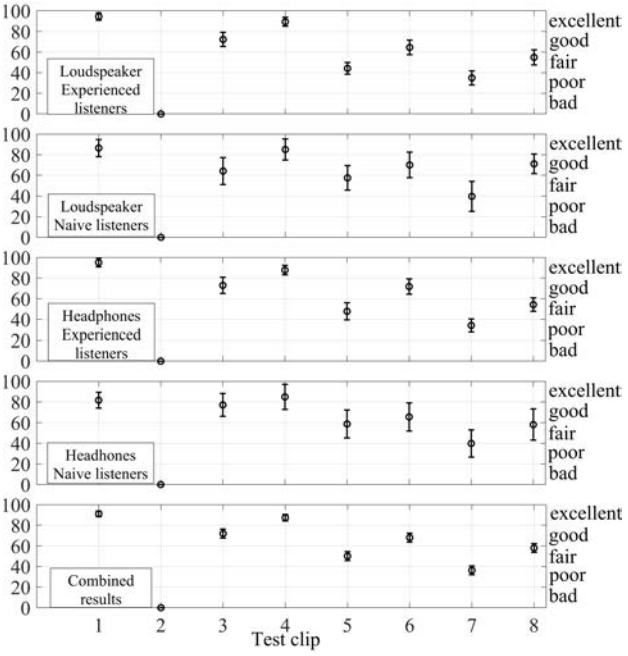

Fig. 14. MUSHRA test results. Mean scores shown with 95% confidence intervals for six different test materials and high/low anchors, numbered as in Table 2

processing scored higher than their non-transient detected counterparts.

To assess if there was a statistically significant difference between the performance of the different source materials, a two-way ANOVA with replication was performed with the null hypothesis that different source materials would have no significant effect. With significance threshold of P-value ≤ 0.05, the P-value was 0.12, supporting the null hypothesis that different source materials had no significant effect on the results.

Fig. 14 breaks down the results by listener experience and sound reproduction method. Another two-way ANOVA with replication found that there was no significant effect of sound reproduction method on the results, with significance threshold was set at P-value ≤ 0.05 and actual P-value of 0.79.

Due to the difference in the number of experienced and naïve listeners, a two-way ANOVA with repetition was not performed to establish the presence of any statistically significant difference between the scores of the two groups. However, there is a strong similarity in the results of the two groups. The same trends are observed, with significant overlap of 95% confidence interval bars for corresponding results. This indicates that the scores given were largely independent of previous audio subjective test experience.

The results enforce the importance of transient detection in decorrelation algorithms that has also been discussed in [7, 18, 27]. This is especially interesting given the relatively small impact of adding transient extraction to DiSP effectiveness, as shown by Fig. 10.

The high-frequency time constant selection of 3.7 ms, which was obtained in the authors' prior subjective assessment of static DiSP [24], combined with an interpolation factor of 20 with transient extraction gained a rating of

"good" or "fair" in all cases. This suggests this level of processing may be used, depending on the requirements of the user, as it will provide a greater level of source decorrelation than the 1 ms/30 interpolation factor level of processing.

Importantly, the high-frequency time constant selection of 1 ms with interpolation factor of 30 with transient extraction scores "excellent" in all cases. This indicates that dynamic DiSP can be applied in a perceptually-transparent manner.

Further work in the form of simulations and real-world case studies needs to be done to assess the levels of decorrelation performance achieved by TDIs with these generation parameters in a variety of scenarios. Specifically, target performance needs to be defined. The results presented in Sec. 4 give a rough idea of the performance of TDIs generated with these parameters, however the data gained only gives information about that particular system. There may need to be further optimization to maximize the level of decorrelation to achieve a target response for a given application. Primarily this would include either a "quality control" stage in the TDI library generation process, where overall TDI library correlation is not allowed to exceed a certain threshold, or data analysis of a large number of generated libraries to isolate TDI combinations that perform optimally.

6 ALTERNATIVE APPLICATIONS

In this work the effectiveness of dynamic DiSP has been examined with particular reference to low-frequency decorrelation. The processing is applicable to live sound reinforcement subwoofer arrays and small room room-mode suppression. Dynamic DiSP has been shown to work to a degree with only one subwoofer present [25] and cinema B-Chains [31].

One of the key benefits of dynamic DiSP is that the solution to the problems caused by coherent source and reflection interference exists within the signal chain as opposed to other measurement-based methods of correction, making dynamic DiSP straightforward and quick to implement.

The TDIs generated with the described parameters provide decorrelation up to Nyquist frequency if necessary, but decorrelation can alternatively be constrained to a specific frequency band by setting the decay time constants of all frequencies outside the band of interest to an arbitrarily small decay time (<1 ms). As previously described, this has the effect of decreasing noise tail amplitude in relation to the amplitude of the impulse, so some post-generation re-scaling of the noise tail amplitude may be necessary to ensure adequate decorrelation. Further subjective tests are necessary to quantify limits for this.

To illustrate the flexibility of DiSP, Fig. 15 shows three TDI magnitude responses—one for full band decorrelation generated with the decay time constants shown in Table 1 and two others where the same decay constants are used apart from setting those frequencies outside the band to 0.1 ms decay constant.

The amplitude and density of spectral notches seen in Fig. 15 are proportional to the level of decorrelation achieved when multiple TDIs with the same generation parameters interact. The random phase of each TDI means the notches appear in different places and with different amplitudes but are constrained to the band in question.

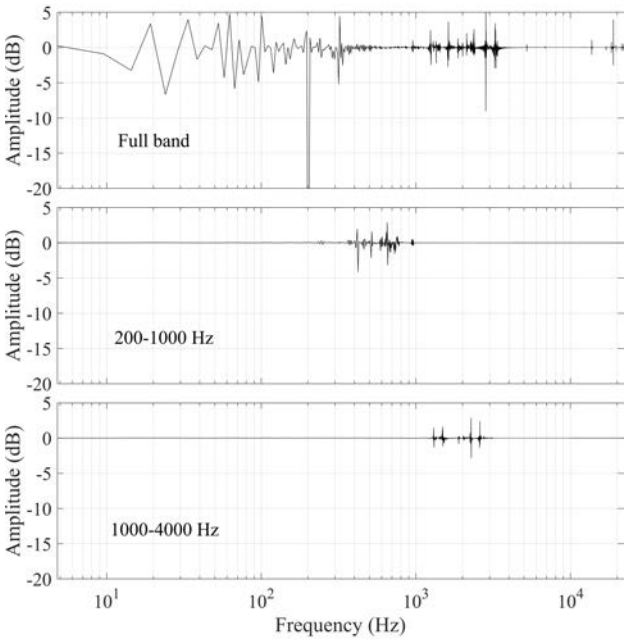

Fig. 15. Impulse magnitude responses of TDIs generated for decorrelation of specific frequency ranges

This means that TDI generation can be tailored to a number of sound reinforcement and reproduction applications. Different elements including subwoofers, main L/R arrays, outfills, frontfills, sidefills, and monitor wedges of PA systems may be decorrelated from each other. Similarly, loudspeakers comprising of two or more drive-units may benefit from TDIs generated to decorrelate around the crossover frequencies. Other applications focus on the improvement of intelligibility of voice PA systems.

7 CONCLUSIONS AND FURTHER WORK

This work describes a time-varying decorrelation algorithm with transient extraction termed *dynamic DiSP*. The effectiveness of the algorithm for the reduction of low-frequency spatial variance in a closed acoustic space has been investigated, and it has been shown that the processing is capable of reducing low-frequency spatial variance in the simulated system by between 25–50%, depending on algorithm settings. Suggestions have been made for the control of the algorithm to be constrained to only three user-controlled parameters: high-frequency decay constant, interpolation factor, and TDI update rate. These parameters give a good deal of flexibility in terms of performance versus perceptibility. Results from a MUSHRA test indicate that the TDI update rate may be set at 15 ms, which is sufficient for all but the smallest rooms, while still retaining "good" or "excellent" audio quality. At the least audible parameter settings (test clip 4), the processing with transient extraction has been shown to be

perceptually transparent, while still providing signal decorrelation as shown in Fig. 9 where a reduction in SV of around 15% was achieved.

There remains an important question: What is a sufficient level of signal decorrelation to obtain the required result for a given application? Currently this can be decided by the user, but further work is necessary to establish clear limits for this. Additionally, the subjective test presented here only focused on processing transparency when compared to an unprocessed sample with one source. Further subjective tests should to be conducted to assess the subjective impact of the dynamic DiSP algorithm and decorrelation when applied to real-world sound reinforcement and reproduction systems.

Another area for further work is improved efficiency of the transient extraction method as the one described here is too slow for real-time processing. Without transient extraction, the only real-time processing in the algorithm is the convolution of each time frame's TDI with a mono source signal, which is computationally inexpensive.

Overall, dynamic DiSP has the potential to provide perceptually-transparent signal decorrelation for a wide-range of sound reinforcement and reproduction applications. The processing can be easily implemented with no system measurements necessary—just a few basic parameters are required. An easily realizable goal is that the user parameters of interpolation factor and high frequency time constant may be controlled by a single user input—a dial, for example, that will give users fine control over level of decorrelation desired versus perceptibility for any application in real-time. Based on this user input, and any frequency limits for decorrelation, the appropriate TDI set for a given time frame can be drawn from a pre-generated suite of TDI libraries allowing for flexible, computationally inexpensive real-time decorrelation.

8 REFERENCES

[1] G. Müller and M. Möser, eds., *Handbook of Engineering Acoustics* (Springer Science and Business Media, 2013). https://doi.org/10.1007/978-3-540-69460-1

[2] A. J. Hill and M. Hawksford, "On the Perceptual Advantage of Stereo Subwoofer Systems in Live Sound Reinforcement," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 8970

[3] A. Celestinos and S. B. Nielsen, "Optimizing Placement and Equalization of Multiple Low Frequency Loudspeakers in Rooms," presented at the *119th Convention of the Audio Engineering Society* (2005 Oct.), convention paper 6545

[4] T. Welti and A. Devantier, "In-Room Low Frequency Optimization," presented at the *115th Convention of the Audio Engineering Society* (2003 Oct.), convention paper 564

[5] D. M Howard and J. A. Angus, *Acoustics and Psychoacoustics*, 4th Ed. (Focal Press, 2009).

[6] H. Lauridsen, "Nogle forsog med forskellige former rum akustik gensivelse," *Ingenioren*, no. 47, p. 906 (1954).

[7] M. Schroeder, "An Artificial Stereophonic Effect Obtained from a Single Audio Signal," *J. Audio Eng. Soc.*, vol 6, pp. 74–79 (1958 Apr.).

[8] R. Orban, "A Rational Technique for Synthesizing Pseudo-Stereo from Monophonic Sources," *J. Audio Eng. Soc.*, vol 18, pp. 157–164 (1970 Apr.).

[9] M. Gerzon, "Signal Processing for Simulating Realistic Stereo Images," presented at the *93rd Convention of the Audio Engineering Society* (1992 Oct.), convention paper 3423.

[10] M. Wilde, W. Martens, and G. Kendall, "Method and apparatus for creating decorrelated audio output signals and audio recordings made thereby," US patent 5235656 (August 10, 1993).

[11] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering for Signal Decorrelation," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 6 (1998 May). https://doi.org/10.1109/ICASSP.1998.679684

[12] D. Morgan, J. Hall, and 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 (2001 Sep.). https://doi.org/10.1109/89.943346

[13] 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 (2004 Oct.).

[14] G. Kendall, "The Effects of Multichannel Signal Decorrelation in Audio Reproduction," *ICMC Proceedings*, pp. 319–326 (1994).

[15] M. Bouéri and C. Kyirakakis, " Audio Signal Decorrelation Based on a Critical Band Approach," presented at the *117th Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6291.

[16] E. Kermit-Cranfield and J. Abel, "Signal Decorrelation Using Perceptually Informed All-Pass Filters," *Proceedings of the 19th International Conference on Digital Audio Effects*, pp. 225–231 (2016 Sep.).

[17] R. Penniman, "A General Purpose Decorrelation Algorithm with Transient Fidelity," presented at the *137th Convention of the Audio Engineering Society* (2014 Oct.), convention paper 9170.

[18] B. D. Molle, J. Pinkl, and M. Blewett, "Decorrelated Audio Imaging in Radial Virtual Reality Environments," presented at the *139th Convention of the Audio Engineering Society* (2015 Oct.), e-brief 208.

[19] V. Pulkki, J. Merimaa, and T. Loki, "Reproduction of Reverberation with Spatial Impulse Response Rendering," presented at the *116th Convention of the Audio Engineering Society* (2004 May), convention paper 6057.

[20] V. Pulkki, J. Merimaa, and T. Loki, "Multichannel Reproduction of Measured Room Responses," International Congress on Acoustics, pp. 1273–1276 (2004).

[21] V. Pulkki, "Spatial Sound Reproduction with Directional Audio Coding," *J. Audio Eng. Soc.*, vol. 55, pp. 503–516 (2007 June).

[22] M. O. J. Hawksford and N. Harris, "Diffuse Signal Processing and Acoustic Source Characterization for Applications in Synthetic Loudspeaker Arrays," presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), convention paper 5612.

[23] J. B. Moore and A. J. Hill, "Optimization of Temporally Diffuse Impulses for Decorrelation of Multiple Discrete Loudspeakers," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), convention paper 9794.

[24] J. B. Moore and A. J. Hill, "Dynamic Diffuse Signal Processing for Low-Frequency Spatial Variance Minimization across Wide Audience Areas," presented at the *143rd Convention of the Audio Engineering Society* (2017 Oct.), convention paper 9903.

[25] J. B. Moore and A. J. Hill, "Applications of Dynamic Diffuse Signal Processing in Sound Reinforcement and Reproduction," *Proceedings of the Institute of Acoustics*, vol. 39, no. 1 (2017 Nov.).

[26] J. B. Allen and D. A. Berkley. "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4 (1979 Apr.). https://doi.org/10.1121/1.382599

[27] A. J. Hill and M. O. J. Hawksford, "A Hybrid Virtual Bass System for Optimized Steady-State and Transient Performance," *Proceedings 2nd Computer Science and Electronic Engineering Conference (CEEC)*, Colchester (2010 Sep.). https://doi.org/10.1109/CEEC.2010.5606489

[28] ITU-R, "Method for the subjective assessment of intermediate quality level of audio systems" (2015).

[29] E. Stein and M. Walsh, "System and Method for Variable Decorrelation of Audio Signals." *U.S. Patent No.* 9,264,838 (16 Feb. 2014).

[30] F. E. Toole, "The Measurement and Calibration of Sound Reproducing Systems," *J. Audio Eng. Soc.*, vol. 63, pp. 512–541 (2015 Jul./Aug.). https://doi.org/10.17743/jaes.2015.0064

[31] A. J. Hill, M. O. J. Hawksford, and P. Newell, "Enhanced Wide-Area Low-Frequency Sound Reproduction in Cinemas: Effective and Practical Alternatives to Current Calibration Strategies." *J. Audio Eng. Soc.*, vol. 64, pp. 280–298 (2016 May). https://doi.org/10.17743/jaes.2016.0012.

THE AUTHORS

Jonathan B. Moore

Adam J. Hill

Jonathan Moore is a Ph.D. research student in the Department of Electronics, Computing and Mathematics at the University of Derby. He obtained a M.Sc. with Distinction in audio production at The University of Salford in 2016. His Ph.D. project is titled "Diffuse Signal Processing for Sound Reinforcement." It focuses on the development of a perceptually transparent audio signal de-correlation algorithm for the reduction of comb-filtering effects in sound reinforcement. Jonathan is a member of the AES and IOA.

•

Adam Hill is currently a senior lecturer at the University of Derby where he runs the M.Sc. Audio Engineering program. He received a Ph.D. from the University of Essex, an M.Sc. with Distinction in acoustics and music technology from the University of Edinburgh, and a B.S.E. in electrical engineering from Miami University. His research generally focuses on analysis, modeling, and wide-area spatiotemporal control of low-frequency sound reproduction and reinforcement. Adam also works seasonally as a live sound engineer for Gand Concert Sound, where he has designed and operated sound systems for over one thousand artists. Adam is chair of the AES Technical Committee on Acoustics and Sound Reinforcement and Head of Content for the Electro-Acoustics Group Committee of the Institute of Acoustics. He is a member of the AES, IEEE, IOA, and IET.