On the use of dynamically varied loudspeaker spacing in wave field synthesis

Ranjan, Rishabh; Gan, Woon-Seng

2012

Ranjan, R., & Gan, W.-S. (2012). On the use of dynamically varied loudspeaker spacing in wave field synthesis. 133rd Audio Engineering Society Convention 2012.

https://hdl.handle.net/10356/106150

Audio Engineering Society
Convention Paper 8755
Presented at the 133rd Convention
2012 October 26–29 San Francisco, CA, USA

On the use of Dynamically Varied Loudspeaker Spacing in Wave Field Synthesis

Rishabh Ranjan and Woon-Seng Gan

DSP Lab, School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, 639798
{rishabh001, ewsgan}@ntu.edu.sg

ABSTRACT

Wave field synthesis (WFS) has evolved as a promising spatial audio rendering technique in recent years and has been widely accepted as the optimal way of sound reproduction technique. Suppressing spatial aliasing artifacts and accurate reproduction of sound field remain the focal points of research in WFS over the recent years. The use of optimum loudspeaker configuration is necessary to achieve perceptually correct sound field in the listening space. In this paper, we analyze the performance of dynamically modified loudspeaker array whose spacing varies with the audio signal frequency content. The proposed technique optimizes the usage of a prearranged set of loudspeaker array to avoid spatial aliasing at relatively low frequencies as compared to uniformly fixed array spacing in conventional WFS setup.

1. INTRODUCTION

The aim of spatial audio system is to create a fully immersive environment for the listener. The two most widely used sound reproduction techniques are binaural reproduction and loudspeaker stereophony, which relies on psychoacoustic principles. Both of these techniques suffer from inherent limitations. Binaural technology suffers from front-back confusion, in head localization, and incorrect perception of virtual source elevation while in multichannel stereophonic reproduction, we can feel the presence of sweet spot, and phantom source perception on the lines connecting the loudspeakers. In the early nineties, Berkhout [1] proposed a novel technique called wave field synthesis (WFS), which relies on the holophonic principles that retain the physical properties (temporal and spatial) of sound fields in the listening area. The aim of WFS based reproduction system is to create a replica of the true sound field in a large reproduction area with effectively no “sweet spot” and enables perfect virtual source localization. In theory, WFS can replicate the original sound source spectrum in the entire listening area but practical approximation of WFS suffers from artifacts, such as amplitude errors, truncation effects, and spatial aliasing errors [1, 2, 3]. Out of these, spatial aliasing is considered to be the most critical as the reproduced sound field is only correct up to aliasing frequency corresponding to the maximum loudspeaker spacing. Spatial aliasing introduces physical inaccuracies in the reproduced sound field at higher frequencies and degrades the perceptual sound quality to a limited extent. One can observe aliasing artifacts as reverberation effect [4], as well as instability in virtual source localization [5]. Timbre changes in the synthesized sound can be perceived if loudspeaker spacing is significantly more than that defined by aliasing criterion [6]. High quality sound reproduction in WFS comes at the cost of umpteen closely spaced loudspeakers, thereby increasing the computational complexity.

In a practical implementation of WFS system using linear array of loudspeakers, uniform spacing (around 10 to 20 cm apart) is typically deployed to avoid the spatial aliasing below 1500 Hz [7], as interaural time difference (ITD) cues are dominant below this frequency [8]. Thus, accurate sound field synthesis below 1500 Hz ensures sound localization in azimuthal (listener) plane. Several methods have been proposed in the literature to suppress the aliasing artifacts [7, 9, 10]. Though, loudspeaker spacing plays a decisive role in determining the fidelity of reproduced sound field, it has been found that spatial aliasing artifacts also depend on other parameters, such as source position, listener position and finite length of the loudspeaker array. Recent efforts have been made to optimally design the loudspeaker array configuration by manipulating different parameters, like array configuration (linear, rectangular, bent, circular etc.), spacing, array length, etc. to improve the performance of the WFS based reproduction system. It has been shown that, a significant increase in aliasing frequency can be achieved with the use of logarithmically spaced array for non-focused sources [11]. Sound reproduction in a “preferred listening-zone” leads to substantial increase in spatial aliasing frequency, in addition to drastically reducing the number of loudspeakers [12]. Spors and Ahrens [13] gave further insights on spatial aliasing using wavenumber domain analysis by deriving the spatial aliasing criterion that also includes secondary source characteristics. Furthermore, in [14], a novel WFS approach using compressive sampling is proposed. This latter approach outperforms their least square counterpart in accurate reproduction of sound field, as well as enabling judicious placement of a minimal number of loudspeakers from a dense grid of loudspeakers. The main objective behind such approaches is to modify the driving signals (loudspeaker signals) so as to reduce the adverse effects of aliasing or to increase the aliasing frequency for target listening area.

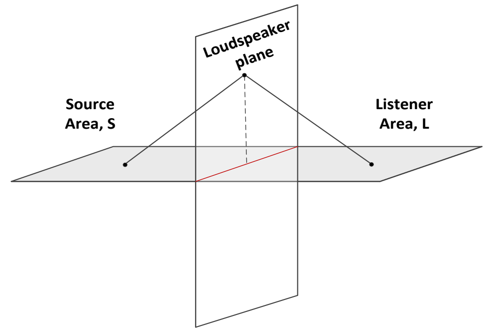

Figure 1 Geometry for Rayleigh Integral

In this paper, we investigate the use of dynamically varied loudspeaker spacing by modifying the driving signals based on frequency content of the source signal. The proposed technique minimizes the loudspeaker usage for a pre-arranged set of loudspeaker array, while avoiding the spatial aliasing at relatively lower frequencies when compared to conventional WFS with uniformly fixed-spaced loudspeaker array.

2. WFS: BACKGROUND AND THEORY

The basic principle of WFS is based on the Huygens principle and mathematically defined by the Kirchhoff Helmholtz Integral (KIH) [1]. The KIH states that sound pressure at any point in an enclosed volume can be calculated if we know the pressure and pressure gradient due to primary source at the surface of source free enclosed volume. In other words, infinite numbers of monopole and dipole secondary sources reproduce the desired sound field inside the volume, while cancelling sound waves outside the volume.

2.1. Practical Approximations for WFS

For practical realization of WFS, Rayleigh showed that surface of enclosed volume can be degenerated into an infinite plane covered by loudspeakers and thus, introduced two Rayleigh integral I and II by eliminating either monopole or dipole term in the KIH integral, respectively. Rayleigh integrals also form the basis for driving signals in WFS. Figure 1 shows an infinite plane of loudspeakers separating the source area, S from the listener area, L. The Rayleigh integral I, from which the driving signals are derived, states that the sound

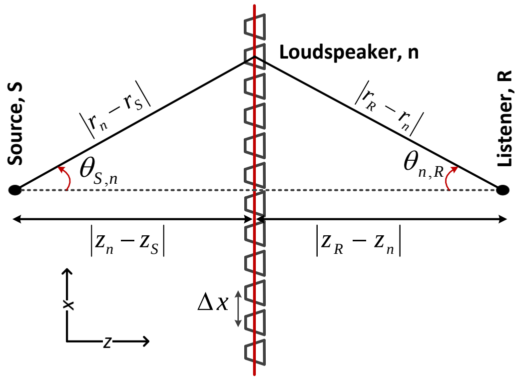

Figure 2 Geometry for sound field synthesis in WFS

pressure in the listening area (L) is recreated from primary or virtual source(s) in the source area (S), with monopole sources as secondary sources on the infinite plane [1, 4]. Elimination of dipole sources from KIH results in non-zero sound field outside the listening volume and inaccuracies inside the listening area due to interferences caused by reflections from undesired loudspeakers [15]. A window function is thus used to suppress the artifacts, which selects only those secondary sources that contribute to sound reproduction.

Other approximations, which are essential in practical realization of WFS, include

- Planar to linear array reduction: Due to the infeasibility of installing many loudspeakers over the plane, we can restrict the number of loudspeakers to linear array.

- Infinite continuous array to finite discrete array: Secondary sources are replaced by fixed finite number of loudspeakers.

The two-dimensional form of Rayleigh integral I is derived as

$$P(R,w) = \frac{1}{4\pi}\sum_{n=1}^{N} D_{WFS}(n,w)\frac{e^{-j\frac{w}{c}|r_n - r_R|}}{|r_n - r_R|}\Delta x, \quad (1)$$

where P is the synthesized sound pressure at point R in the listening area, L (Figure 1,2). r_n and r_R are the position vectors of the n^{th} loudspeaker and the listener position, R, respectively. The notation D_{WFS} represents the driving function for loudspeaker n and is the principal function of a WFS system. N is the total number of loudspeakers for a given length of array. The geometry for (1) is shown in figure 2. The exponential part in (1) is the three dimensional Green's function, which represents the radiation of a secondary monopole point source [16]. In time domain, sound pressure at any listener point can be simply calculated by summing the weighted and delayed contribution from each loudspeaker.

2.2. Driving Signals

The solution for driving signal at n^{th} loudspeaker is obtained by equating (1) with sound field of a monopole point source in free-field at listener position (R), using stationary phase approximation [2] as

$$D_{WFS}(n,w) = S(w)\sqrt{\frac{jw}{2\pi}}\sqrt{\frac{|z_R - z_n|}{|z_R - z_S|}}\cos(\theta_{S,n})\frac{e^{-j\frac{w}{c}|r_n - r_S|}}{\sqrt{|r_n - r_S|}}. \quad (2)$$

Thus, driving signal at each loudspeaker can be simply calculated by summing the delayed and weighted contribution from filtered source(s) signal(s).

The source signal is first pre-filtered to compensate for the planar-to-linear array approximation. Third term in (2) is the correction factor and also accounts for the 2D approximation. As evident, it depends on the source-receiver distance and source-listener distance along the z-plane. Thus, the driving signal is weakly dependent on receiver position (in fact on the receiver line, Z_R) and as a result synthesized sound pressure field is correct only on the reference receiver line, which is used in the computation of driving signal. The second last term in (2) is a weighting factor depending on the source incidence angle as shown in Figure 2. The last term represents the natural wave propagation and amplitude decay of source, S.

2.3. Practical Limitations: Effects and Solutions

As discussed earlier, approximations employed to KIH integral for practical realizations of WFS put constraints on the synthesized sound field quality, visible area, number of loudspeakers, array length, and array configuration.

The planar-to-linear array reduction introduces inaccuracies in the sound pressure attenuation, which is compensated by pre-filter having 3dB per octave attenuation and a phase shift of 45^o [12]. This approximation results in perfect sound field reproduction on the azimuthal (listener) plane but source localization perception in vertical plane is inaccurate. Multiple linear arrays in vertical plane can be employed to improve the elevation perception problem [17].

Reduction to finite length array due to practical infeasibility introduces truncation effects in the reproduced sound field, which are observed as circular trailing waves originating from the extremes of the loudspeaker array. Truncation effects (also known as diffraction effects) can be suppressed by tapering the extreme loudspeakers, but at the cost of reduced effective array length. Effective array length can be further increased by extending the loudspeaker array on side walls of the listening room and tapering applied at the two extremes [2]. In addition, finite length of the array reduces the visible area for listener, which is defined by extent of the loudspeaker array.

In practical situation, secondary sources are replaced by discrete number of loudspeakers. Thus, continuous linear array is reduced to discrete array by applying spatial sampling and suffers from spatial aliasing similar to that in temporal domain sampling. In this context, loudspeaker spacing plays an important part in synthesizing desired sound field. Sound field can be faithfully reproduced if frequency contents of the source are below the spatial aliasing frequency else distortions in frequency response, physical sound field, and perceptual quality can be observed. Spatial aliasing effects can be avoided by several methods, such as spatial bandwidth reduction by means of secondary source directivity characteristics to reduce the interference [3, 7] and optimized phantom source imaging (OPSI) by reproducing high frequency content via few selected loudspeakers acting as phantom source [6].

With large number of closely-spaced loudspeakers and a general sound scene with multiple sound sources, WFS is computationally complex as well as power intensive. Loudspeaker usage has to be in check for practical realization of WFS in real time applications. Therefore, it is necessary to use optimum array configuration, which are mostly governed by loudspeaker spacing and array configuration, to warrant the perceptually correct sound field in target listening area. Spatial aliasing criterion is analyzed in detail in the next section with a focus on selecting the best aliasing criterion.

3. OPTIMUM SPATIAL ALIASING CRITERION

Prominent distortions are observed in the synthesized sound field as a result of spatial aliasing. Spatial aliasing effects can be better understood with the help of aliasing criterion, which is expressed in terms of source, loudspeaker and listener parameters. Spatial aliasing criterion is usually defined as the frequency above which spatial aliasing effects are bound to occur.

There are several mathematical definitions of spatial aliasing criterion in the literature mainly because of assumptions on varied practical scenarios. There are mainly four factors on which spatial aliasing depend:

1. Loudspeaker spacing
2. Source position / Source orientation / Source incidence angle
3. Listener position / Loudspeaker incidence angle at listener position
4. Loudspeaker array length / Number of loudspeakers

Since linear array is spatially sampled with loudspeaker spacing, it is the key factor in determining the aliasing frequency. Other factors are also crucial under given circumstances. Table 1 shows the expression for 5 different spatial aliasing criterions available in the literature. Third column in the table represents the list of factors on which each criterion depends.

Spal1 is the most relaxed aliasing criterion and considered as the worst case criterion in any scenario [1]. Expression for **Spal1** is derived by applying sampling theory which states that for exact reproduction of sound field, sampled spacing should be less than or equal to half the wavelength as given in Table 1. It is clear from the expression that higher the frequency of the source signal, closer the loudspeakers should be placed to avoid aliasing and vice-versa.

Spal2 is a more constrained criterion with respect to source orientation. It is derived in wavenumber domain and plane wave decomposition using the notion that apparent wavelength component along the linear array should be greater than or equal to loudspeaker spacing [7]. This criterion depends on loudspeaker spacing as well as incidence angle of the plane wave source with linear array. We can clearly observe from the expression in Table 1, that when plane wave is perpendicular to the linear array, the criterion is same as **Spal1.**

Table 1 List of several definitions of spatial aliasing criterion available in the literature

Spatial Aliasing Criterion	Expression	Dependency Parameters
Spal1 [1]	$\frac{c}{2\Delta x}$	Loudspeaker spacing
Spal2 [7]	$\frac{c}{2\Delta x \sin\theta_{S,pw}}$	Loudspeaker spacing, source orientation (plane wave source)
Spal3 [18, 19]	$\frac{c}{\Delta x(1+\|\sin\theta_{S,pw}\|)}$	Loudspeaker spacing, source orientation (plane wave source)
Spal4 [10]	$\frac{c}{\Delta x(\|\sin\max(\theta_{S,n}) - \sin\max(\theta_{n,R})\|)}$	Loudspeaker spacing, source orientation, loudspeaker incidence angle, array length
Spal5 [11, 12]	$\frac{c}{\Delta x \max_n(\|\sin\theta_{S,n} - \sin\theta_{n,R}\|)}$	Loudspeaker spacing, source orientation, loudspeaker incidence angle, array length

Spal3 is derived using an approach similar to temporal domain sampling for plane wave sources [18] and has also been extended to broadband signals [19]. The difference from **Spal2** is due to the fact that it also includes the radiation characteristics of secondary sources in the derivation. This criterion is thus more suited for practical set ups.

Spal4 and **Spal5** depend on the source and listener positions as well as loudspeaker array length. Both these criterions take care of the fact that perceived aliasing artifacts vary with the listener and source movement. Aliasing frequency depends on maximum source incidence angle in accordance to spatial sampling of the loudspeaker array as mentioned in [20]. Likewise, for exact reconstruction, aliasing frequency also depends on the maximum of loudspeaker incidence angles at listener position [10], which is the basis for criterion **Spal4** as shown in Table 1. A similar explanation is also given in [12]. Aliasing frequency can be increased by increasing the directivity of source and loudspeaker.

Spal5 is derived using temporal domain analysis according to which aliasing frequency can be expressed as a function of arrival time difference between contributions of consecutive loudspeakers at listener position and can be approximated as shown in Table 1 [11, 12]. In contrast to **Spal4,** this criterion considers the relative position of source and listener and thus accurately matches with behavior of aliasing artifacts with frequency in practice. It should be noted that aliasing frequency will be maximum if the source, loudspeaker and listener are all aligned, i.e. they are in stationary phase. Next we validate the aliasing criterion **Spal5** with the simulated frequency response of loudspeaker array. We can easily observe that **Spal1** is special case for all the criterions.

In this paper, we focus on linear loudspeaker array to show our approach. Figure 3 shows the WFS simulation set up with a linear loudspeaker array of 4m uniformly spaced at 2.5 cm. There are five possible positions for virtual source as shown, i.e. from S1 (-4,-10) to S5 (4,-10) each 2m apart. Listener positions are fixed on a line parallel to array and at a distance of 1m from the array. Listener positions are 0.2m apart and 2m in length.

Spatial aliasing in WFS leads to distortion in frequency response of the loudspeaker array and thus deviates from the flat frequency response above aliasing frequency as shown in Figure 4 (a), (b) and (c). Frequency distortion can be observed as sharp peaks and dips in the frequency response, which is otherwise approximately flat below certain frequency. These distortions are nothing but the spatial aliasing artifacts.

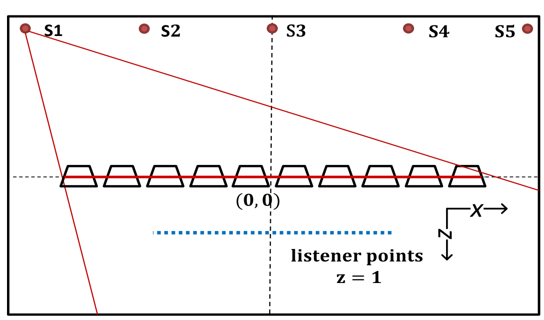

Figure 3 WFS simulation set up with 161 loudspeakers uniformly spaced at 2.5 cm

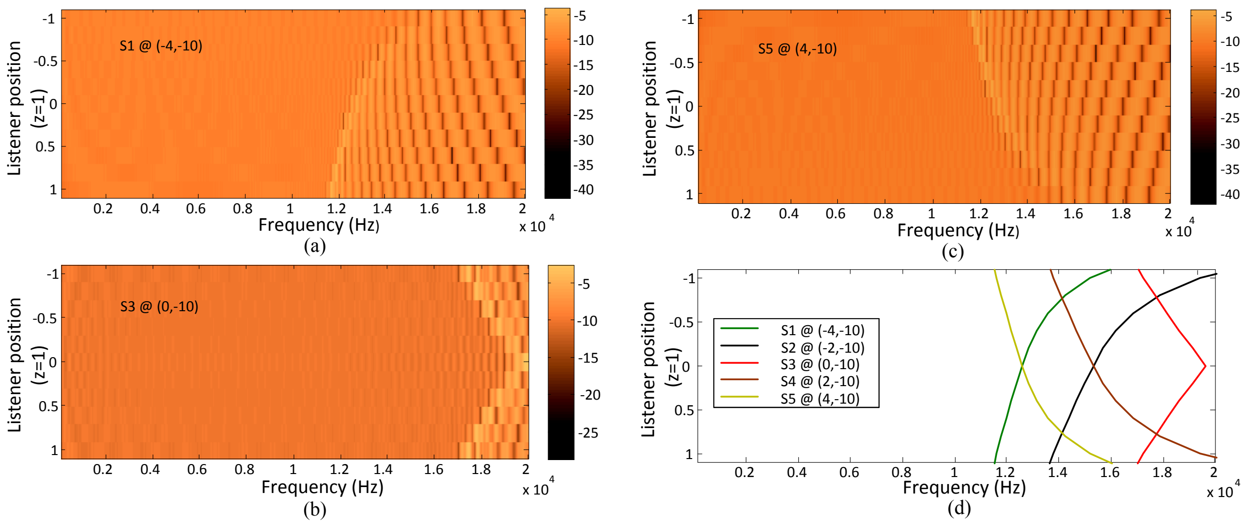

Figure 4 Spatial aliasing criterion validation with simulated frequency response of loudspeaker array in WFS: (a), (b), (c) Frequency response of array with virtual source located at different positions (d) Spatial aliasing frequency for all listening positions and virtual source positions using criterion **Spal5**

The variation in aliasing effect with source and listener position can be clearly discerned. Validity of the **Spal5** aliasing criterion can be confirmed by comparing simulated frequency response in Figure 4 (a), (b) and (c) with aliasing frequency plots using **Spal5** in Figure 4(d) for respective source positions. It is noted that aliasing artifact is severe when source and listener are far apart and minimum when both are aligned with loudspeaker in stationary phase**.** Thus, this criterion can be chosen as the optimum aliasing criterion and is used for our further analysis in the paper.

4. WAVE FIELD SYNTHESIS USING DYNAMICALLY VARIED LOUDSPEAKER ARRAY

Fidelity of sound field synthesis in WFS is primarily dependent on frequency of the source signal. In a practical WFS set up, we generally pre-arrange the loudspeakers position around the extended listening area. Thus sound fields are correctly synthesized below aliasing frequency but at the cost of high energy usage since all the loudspeakers are active all the time. A dynamic decision on selecting the loudspeakers weights at regular time intervals can reduce the loudspeaker usage for a given arrangement. In this paper, we present a technique to construct a dynamically spaced loudspeaker array using short-time Fourier transform (STFT) analysis based on an optimal loudspeaker spacing selection algorithm. Our approach exploits the fact that most of the real audio signals are non-stationary and possess wide frequency spectrum. Frequency content of the audio signal can thus be used to make a decision on the optimal spacing so as to avoid spatial aliasing to an extent, while keeping the number of active loudspeakers to a minimum.

4.1. Methodology

The strategy of our method is to use the frequency bandwidth of audio signal to optimize the loudspeaker usage by allocating weights to loudspeakers in real time. Since spatial aliasing frequency is inversely proportional to loudspeaker spacing, highest frequency component of the source spectrum is used to decide on the optimum loudspeaker spacing to select a co-array from densely spaced array subject to aliasing constraint. Figure 5 shows the flow diagram for our proposed method using WFS, which dynamically select a co-array with uniform spacing as integer multiple of minimum spacing of pre-arranged array. Frame size is chosen sufficiently large to allow for smooth switching of loudspeakers in real time. Source position, listener position, array parameters along with the pre-defined set of loudspeaker spacings is given as input to system.

4.1.1. Maximum frequency estimation and LUT formation

Audio input is processed frame by frame using overlap add (OLA) method. Using STFT analysis, we estimate the maximum frequency, f_{max}, in two steps: first, forming a set of frequencies corresponding to all the local maximas of magnitude power spectrum of the

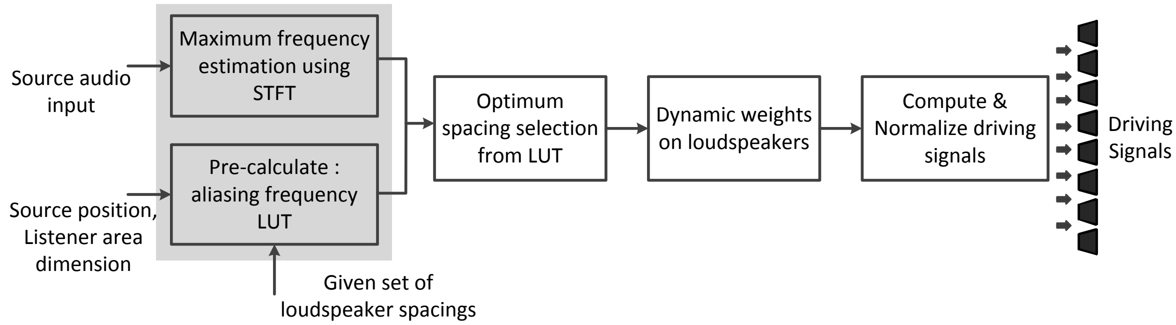

Figure 5 Flow diagram showing procedure of dynamically varied loudspeaker array in Wave Field Synthesis

source signal, and second, by discarding those frequency components for which magnitude of the peaks are less than a threshold energy value and then, taking maximum of remaining frequencies in the set. Thus, the threshold value can be tuned to have better estimate of source spectrum. It is obvious that larger the threshold value, higher frequency components will be estimated and subsequently, denser co-array will be selected leading to less erroneous reproduced sound field at the cost of high loudspeaker usage and vice-versa. Using the aliasing criterion **Spal5,** we also pre-calculate the aliasing frequencies and store them in a look up table (LUT), f_{al}, for the whole set, S, of conceivable loudspeaker spacings (see fig. 5), given source position and listener positions.

4.1.2. Co-array selection and dynamic weights to loudspeakers

We choose the co-array from pre-arranged densely spaced loudspeaker array by estimating the aliasing frequency based on estimated maximum frequency and pre-calculated LUT as

$$f_{al,j}^{estimated} = \{\min(f_j \in f_{al} > f_{\max} \pm \delta f) \quad \forall \quad j\}, \tag{3a}$$

$$\Delta x \;=\; S(j), \tag{3b}$$

where Δx is the selected optimal co-array spacing for current frame such that the estimated aliasing frequency is minimum. δf is the small value (within ~10% of the frequency) which accounts for the slight deviation of **Spal5** from real values of aliasing frequency (Figure 4). Once the co-array is selected, loudspeakers are dynamically allocated weights according to their significance at each frame. Loudspeakers, which belong to the selected co-array, are essential for aliasing free sound reproduction and thus are given full weight (value of 1 on the scale of 0-1) while rest loudspeakers are assigned non-zero but small weights. Thus, dynamic selection window for the nth driving signal is defined as

$$W(n) = \begin{cases} 1, & (n-1)\%\left(\dfrac{\Delta x}{d}\right) == 0 \\ \mu, & (n-1)\%\left(\dfrac{\Delta x}{d}\right) \neq 0 \quad ; \mu << 1 \end{cases}, \tag{4}$$

where μ is the non-zero but small weights assigned to loudspeakers and d is the spacing of densely spaced pre-arranged array. Non-zero weights to loudspeakers increase the spatial aliasing frequency while reducing the distortions in low frequencies. Fading window is also applied to driving signals to account for the spectral leakage due to real-time switching of loudspeakers.

4.1.3. Modified Driving Signals

Modified driving signal can thus be expressed as

$$d_{WFS}^{\text{modified}}(n,t,k) = \frac{d_{WFS}(n,t,k) \times W(n,k)}{eq(k)}, \tag{5a}$$

$$eq(k) \;=\; ls(k) + (N - ls(k)) \times \mu, \tag{5b}$$

where $d_{WFS}(n,t,k)$ is the conventional WFS driving signal (2) in time domain for n^{th} loudspeaker and at k^{th} frame. Driving signal is thus multiplied by selection window and normalized by the term *eq(k)*, which represents effective number of active loudspeakers at frame, *k*. This factor compensates for uneven number of active loudspeakers at successive frames resulting in loudness variation. *ls(k)* is the number of loudspeakers in the selected co-array at frame *k*.

5. SIMULATION RESULTS

As shown in Figure 3, we consider here a densely spaced linear array of 4m in length with uniform spacing of 2.5cm and listener positions of 2m in length.

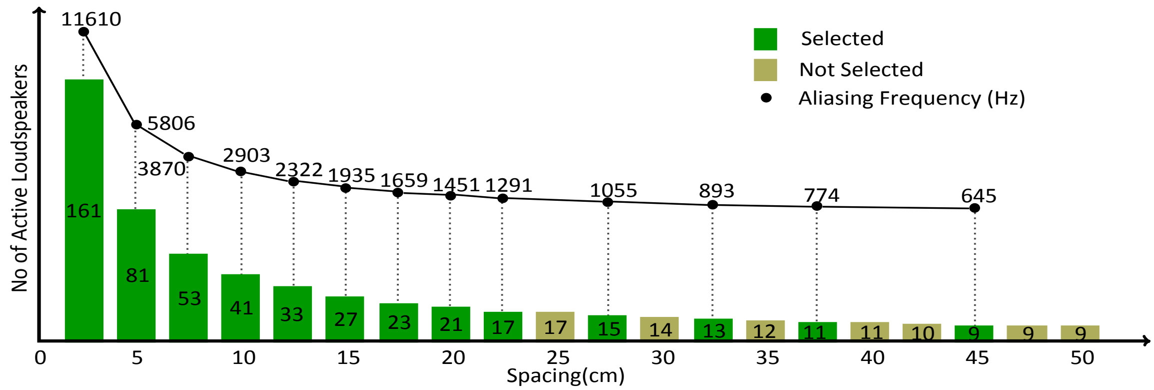

Figure 6 Selection of co-array spacing and associated aliasing frequency

We form the set, S, of loudspeaker spacings (as integer multiple of minimum spacing i.e. 2.5cm). Figure 6 shows the plot of loudspeaker usage and aliasing frequency vs. loudspeaker spacing (ranging from 2.5cm to 50cm). Only those spacing is used in the set, S for which number of active loudspeakers reduces at least by two from the left adjacent spacing. Thus, from Figure 6, we have in total 13 spacings (up to 45 cm) available for co-array selection with aliasing frequency varying from 11.6 KHz to as low as 645 Hz. This gives us the comprehensive range of frequencies to apply our method to most of the audio signals.

5.1. Test Signals

Three test signals with different spectrum characteristics are used in the simulations, namely,

- Drum signal: percussions with frequencies up to around 14-16 KHz (positioned at S3, Figure 3)
- Speech signal: vocals with frequencies up to around 3-4 KHz (positioned at S1)
- Sine sweep signal: exponential sine sweep with frequencies from 200Hz-17KHz (positioned at S5)

Threshold energy value is used as control parameter and is varied to achieve a trade-off between acceptable perceived quality and optimum loudspeaker usage. We quantify the acceptable perceived quality by computing the average error frequency response for each of the test signals. Loudspeaker savings percentage is calculated using (5b) by taking average over all time frames. We used the μ value equal to 0.1 for numerical simulations.

5.2. Loudspeaker Savings using Dynamically Modified Array

Table 2 shows the mean error deviations (MED) and loudspeaker savings of proposed method as compared to conventional WFS for 4 threshold energy values (10 to 30 dB). Spectrum of the reproduced sound is divided into two regions, namely, low frequency region (LF), for frequency less than 1500 Hz and high frequency region (HF), for frequency greater than 1500 Hz to clearly distinguish between the artifacts for the two regions. MED is then computed for both the regions by taking average of the absolute difference between proposed method and conventional fixed WFS for all the listener positions. Loudspeaker savings (LS %) for each situation is also reported in the table. Since ITD cues are dominant cues in low frequencies, less error deviations in low frequency is desired for perceptually acceptable sound quality. From Table 2 we can observe that there are large MED values in high frequencies for all the three test signals in comparison to low frequencies. As the threshold value is increased, MED decrease i.e. proposed method approaches fixed WFS with an increase in loudspeaker usage for all the three test signals. Thus, an optimum value for threshold value can be found by making a trade-off between the MED values for LF region and LS %. Both these parameters are strongly dependent on the nature of the audio signals as can be seen in Table 2.

Table 2 shows large reduction in MED value (~3 dB to ~1 dB) and LS % (82% to 20%) for drum signal as threshold value is increased. Speech signal and sine sweep signal possesses relatively less reduction in MED values with loudspeaker savings of at least 64 % and 66 % respectively. This is mainly because of the nature of the frequency spectrum of percussions in the drum signal which has a wide frequency spectrum and thus, requires high loudspeaker usage (or less LS %). For sine sweep signal, there is very little variation in MED value and LS % which is attributed to the fact that frequency spectrum is very narrow in a frame. Thus, depending on the type of audio signal we can vary the threshold energy value to have perceptually correct sound field as

Table 2 Mean error deviations and loudspeaker savings of proposed method compared to conventional WFS

Threshold energy value	Drum			Speech			Sine sweep		
	Mean error (dB)		LS %	Mean error (dB)		LS %	Mean error (dB)		LS %
	LF	HF		LF	HF		LF	HF	
10 dB	2.947	5.383	82.4	1.751	4.230	80.5	2.277	2.328	67.3
20 dB	2.240	4.763	63.3	1.448	3.955	75.1	2.041	2.133	66.9
27 dB	1.560	3.483	38.9	1.085	3.505	69.3	1.910	1.999	66.8
30 dB	1.072	2.250	19.8	0.917	3.163	64.2	1.778	1.843	66.7

well as reasonable savings on loudspeaker usage.

Figure 7 plots the average error spectrum (over all the listener positions) of the reproduced sound field for the proposed method and conventional WFS with respect to spectrum with no aliasing. Results of the proposed method and fixed WFS with 100 % loudspeaker usage are shown in the figure for drum and speech signal. The error values of the proposed method with high threshold values (27 dB and 30 dB) do not differ from fixed WFS by more than 2 dB at low frequencies. MED values in Table 2 are corroborated by the error spectrum pattern in Figure 7. Therefore, we can easily make a safe decision on the choice of optimal threshold value with loudspeaker savings of at least 40-50 % if error (or MED value) at low frequencies is relatively lesser with performance of the proposed method almost similar to conventional WFS. From Table 2 and Figure 7, threshold values of 27, 20, and 30 dB for drum, speech and sine sweep signal respectively with LS % of at least 40, 75, and 66 % respectively can be taken as good trade-off.

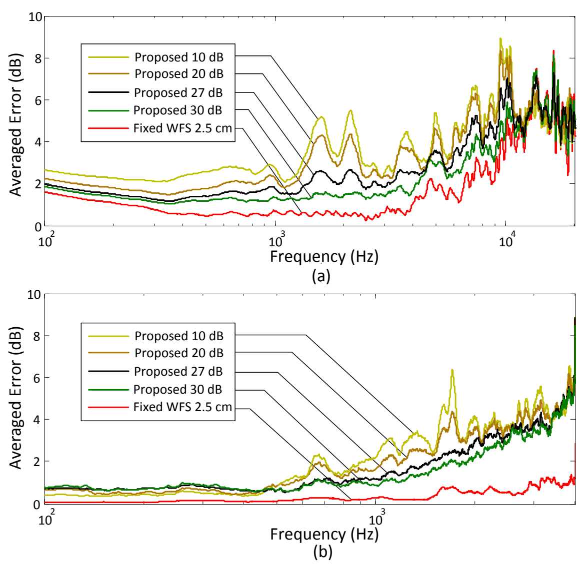

Figure 7 Comparison of average error spectrum for (a) drum signal (b) speech signal

Figure 8 shows the relation between error and frequency of the speech signal for the proposed method and fixed WFS with same effective loudspeaker usage (~40 loudspeakers out of total available 161). Plot for fixed WFS is drawn with spacing 15 cm with same effective number of active loudspeakers. We can visibly notice that there is more error in the frequency range 1.5-3 KHz for conventional WFS while error is uniformly evenly spread out across all the frequencies for the proposed method which is mainly due to dynamic allocation of loudspeaker weights based on source frequency.

6. CONCLUSIONS

In this paper, we presented an approach that dynamically assigns the loudspeaker weights using an optimum co-array selection which reduce the loudspeaker usage by at least 50-60 %. The reproduced sound quality is perceptually correct with very little artifacts in low frequencies warranting correct source localization. Amount of loudspeaker savings as amplifier's energy is heavily dependent on the source signal spectrum characteristics. Our method provides control to adjust the threshold to get the desired response. Our approach can also be combined with other existing techniques to reduce aliasing artifacts and thus present prospects to apply it into real-time applications. Future work entails further experiments and measurements for a general sound scene with moving sound sources as well as different array configurations.

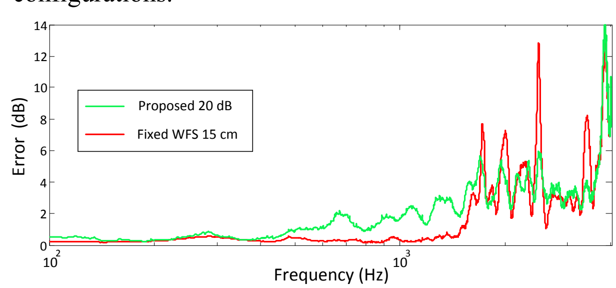

Figure 8 Comparison of error spectrum for speech signal with equal loudspeaker usage

7. ACKNOWLEDGEMENTS

This work was supported by the Singapore Ministry of Education Academic Research Fund Tier-2, under research grant MOE2010-T2-2-040.

8. REFERENCES

[1] Berkhout, A. J., de Vries, D., and Vogel, P., "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764-2778, December 1993.

[2] Vogel, P. *Application of Wave-Field Synthesis in room acoustics*, PhD thesis, Delft University of Technology, May 1993.

[3] Verheijen, E.N.G. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Delft University of Technology, 1997.

[4] de Vries, D. *Binaural Technology* (AES Monograph) Volume 1. Available: http://www.aes.org/publications/monographs/

[5] Sanson J., Corteel E., and Warusfel O., "Objective and subjective analysis of localization accuracy in Wave Field Synthesis," *presented at the 124th Convention of the Audio Eng. Soc.*, Amsterdam, The Netherlands, May 2008.

[6] Wittek, H. *Perceptual differences between wave field synthesis and stereophony*, PhD thesis, University of Surrey, 2007.

[7] Start, E. W. *Direct Sound Enhancement by Wave Field Synthesis*, Ph.D. thesis, TU Delft, Delft, The Netherlands, 1997.

[8] Blauert, J. "*Spatial Hearing: The Psychophysics of Human Sound Localization*," (M.I.T. Press, Cambridge, MA1983)

[9] de Bruijn, W. *Application of Wave Field Synthesis in Videoconferencing*. PhD thesis, TU Delft, Delft, Pays Bas, 2004.

[10] Wittek, H., Rumsey, F., and Theile, G., "Perceptual Enhancement of Wave Field Synthesis by Stereophonic Means," *J. Audio Eng. Soc.*, vol. 55, no. 9, pp. 723-751, 2007.

[11] Corteel, E., "On the use of irregularly spaced loudspeaker arrays for Wave Field Synthesis, potential impact on spatial aliasing frequency," *9th Int. Conference on Digital Audio Effects*, 2006

[12] Corteel, E. and Pellegrini, R. "Wave Field Synthesis with increased aliasing frequency," *presented at the 124th Convention of the Audio Eng. Soc.*, Amsterdam, The Netherlands, May 2008.

[13] Ahrens, J. and Spors, S. "Sound Field Reproduction Using Planar and Linear Arrays of Loudspeakers," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 18, no. 8, pp. 2038–2050, 2010.

[14] Lilis, G. N., Angelosante, D., and Giannakis, G. B., "Parsimonious sound field synthesis using compressive sampling," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2009, pp. 253-256.

[15] Spors, S. Rabenstein, R. and Ahrens, J. "The Theory of Wave Field Synthesis Revisited," *presented at the 124th Convention of the Audio Eng. Soc.*, Amsterdam, The Netherlands, May 2008.

[16] Williams, E. G. *Fourier Acoustics: Sound Radiation and Near-field Acoustical Holography*. Academic Press, 1999.

[17] Montag, M. N. *Wave Field Synthesis In Three Dimensions by Multiple Line Arrays*, MSc. thesis, University of Miami, Florida, 2011.

[18] Spors, S. and Rabenstein, R., "Spatial Aliasing Artifacts Produced by Linear and Circular Loud-speaker Arrays used for Wave Field Synthesis," *presented at the 120th convention of the Audio Eng. Soc.*, Paris, France, May 2006.

[19] Spors S., "Investigation of Spatial Aliasing Artifacts of Wave Field Synthesis in the Temporal Domain," *In Fortschritte der Akustik*, *DAGA*, 2008.

[20] Sonke, J. J., Labeeuw, J., de Vries, D., "Variable Acoustics by Wave Field Synthesis: A Closer Look at Amplitude Effects," *presented at the 104th convention of Audio Eng. Soc.*, Amsterdam, The Netherlands, April 1998.