1-1-2015

An efficient approach to dynamically weighted multizone wideband reproduction of speech soundfields

Jacob Donley
University of Wollongong, jrd089@uowmail.edu.au

Christian H. Ritz
University of Wollongong, critz@uow.edu.au

An efficient approach to dynamically weighted multizone wideband reproduction of speech soundfields

Abstract

This paper proposes and evaluates an efficient approach for practical reproduction of multizone soundfields for speech sources. The reproduction method, based on a previously proposed approach, utilises weighting parameters to control the soundfield reproduced in each zone whilst minimising the number of loudspeakers required. Proposed here is an interpolation scheme for predicting the weighting parameter values of the multizone soundfield model that otherwise requires significant computational effort. It is shown that initial computation time can be reduced by a factor of 1024 with only 85dB of error in the reproduced soundfield relative to reproduction without interpolated weighting parameters. The perceptual impact on the quality of the speech reproduced using the method is also shown to be negligible. By using pre-saved soundfields determined using the proposed approach, practical reproduction of dynamically weighted multizone soundfields of wideband speech could be achieved in real-time.

Keywords

weighted, multizone, wideband, speech, reproduction, soundfields, efficient, approach, dynamically

Disciplines

Engineering | Science and Technology Studies

Publication Details

J. Donley & C. Ritz, "An efficient approach to dynamically weighted multizone wideband reproduction of speech soundfields," in Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on, 2015, pp. 60-64.

AN EFFICIENT APPROACH TO DYNAMICALLY WEIGHTED MULTIZONE WIDEBAND REPRODUCTION OF SPEECH SOUNDFIELDS

Jacob Donley, Christian Ritz

School of Electrical Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia, 2522, jrd089@uowmail.edu.au, critz@uow.edu.au

ABSTRACT

This paper proposes and evaluates an efficient approach for practical reproduction of multizone soundfields for speech sources. The reproduction method, based on a previously proposed approach, utilises weighting parameters to control the soundfield reproduced in each zone whilst minimising the number of loudspeakers required. Proposed here is an interpolation scheme for predicting the weighting parameter values of the multizone soundfield model that otherwise requires significant computational effort. It is shown that initial computation time can be reduced by a factor of 1024 with only $-85dB$ of error in the reproduced soundfield relative to reproduction without interpolated weighting parameters. The perceptual impact on the quality of the speech reproduced using the method is also shown to be negligible. By using pre-saved soundfields determined using the proposed approach, practical reproduction of dynamically weighted multizone soundfields of wideband speech could be achieved in real-time.

Index Terms— multizone soundfield reproduction, wideband multizone soundfield, weighted multizone soundfield, look-up tables (LUT), interpolation, sound field synthesis (SFS)

1. INTRODUCTION

The reproduction of audio in spatially separated regions is of high interest for applications such as the creation of personal listening zones in entertainment/cinema, multi-participant teleconferencing and vehicle cabins. This can be achieved using multizone soundfield reproduction and was originally proposed in [1] to reproduce multiple independent zones of active and quiet 2D soundfield regions. In [1] a least squares pressure matching technique was used, where estimated acoustic pressures reproduced by a set of loudspeakers are matched to sample values within the desired soundfield zone. This technique is also used in Sound Field Reconstruction (SFR) [2].

Following [1], a multizone approach using cylindrical harmonic expansion with coefficient translation and angular windowing was proposed [3]–[5]. This approach, however, attempts to completely suppress any interzone interference which can result in impractically large loudspeaker signal amplitudes or impractically low levels in zones. A method better suited for implementation and controllability was introduced in [6]. The approach uses orthogonal basis expansion which reduces the problem to the reconstruction of a set of basis wavefields and allows each zone to be weighted according to the importance of its reproduction. This weighting improves the practicality of the system by relaxing the ideal requirement of completely quiet zones outside the target zone. The theory in [5] was extended in [7] to include a similar weighting criteria as used in [6].

While originally focusing on single frequency soundfields, more recent work attempted to create multizone soundfields [8], [9] with frequency bandwidths equivalent to narrowband and wideband speech. This paper investigates the use of weighted multizone soundfield reproduction for wideband speech using the orthogonal basis expansion of [6]. The orthogonal basis expansion approach assumes the soundfield is reproduced as a sum of independent planewaves reproduced by an array of loudspeakers. Each planewave corresponds to an individual frequency and direction. The number of chosen planewaves is governed by the desired reproduction accuracy and varies depending on frequency. Due to this, a formula is proposed in this work which determines the number of planewaves to be used for a given frequency so that neither spatial aliasing nor ill-conditioned matrices is an inherent problem. This is necessary when dealing with the reproduction of soundfields containing multiple frequencies.

While [6] assumes the same weight for each frequency, dynamically deriving the weights can be used to control the reproduction accuracy of individual frequency components within the bright and quiet zones. For example, the weightings can be based on the perceptual importance of particular frequencies in the zones in an effort to improve the overall perceived sound quality. However, this results in increases in computational complexity. To reduce this complexity and create a more practical solution, this paper proposes the interpolation of spatial components of the reproduction along different domains, such as the weighting domain and frequency domain.

A system is synthesised with varying linear interpolation distances by using different resolution lookup tables (LUTs) for storing pre-computed loudspeaker weights and soundfield values. The synthesis comprises of reproducing wideband zones where frequency domain content is weighted uniformly with weights that are in the centre of interpolation regions. The approach is validated by comparing the reproduced zone signals from the interpolation method with signals reproduced without interpolation using Mean Squared Error (MSE) and Perceptual Evaluation of Speech Quality (PESQ) [10] measures.

Section 2 describes the orthogonal basis expansion approach to multizone soundfield reproduction. The proposed dynamically weighted multizone approach is described in Section 3 and a novel approach to selecting the number of orthogonal planewaves is presented in Section 4. Section 5 describes the interpolation method. Evaluation results are given in Section 6 and conclusions outlined in Section 7.

2. MONOFREQUENT MULTIZONE SOUNDFIELD REPRODUCTION

A multizone soundfield reproduction layout is depicted in Figure 1. The reproduction region, $\mathbb{D}$, of radius R, contains three sub

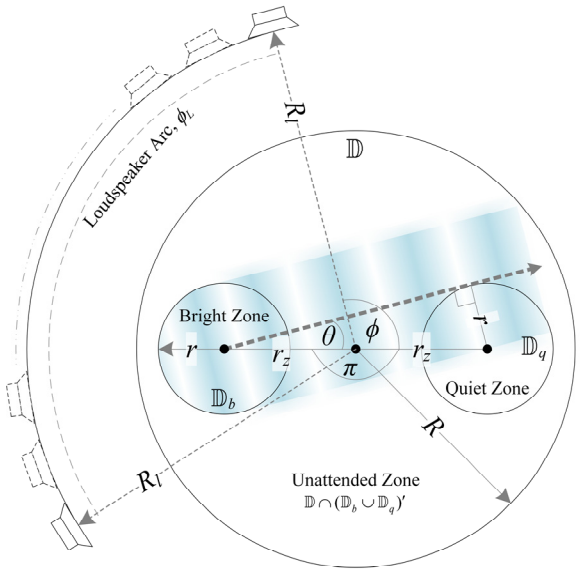

Figure 1 – Weighted multizone soundfield reproduction layout

regions called the bright zone, quiet zone and unattended zone, denoted by $\mathbb{D}_b$, $\mathbb{D}_q$ and $\mathbb{D} \cap (\mathbb{D}_b \cup \mathbb{D}_q)'$, respectively. The radius of $\mathbb{D}_b$ and $\mathbb{D}_q$ is r and their centres are located on a circle of radius r_z. The angle of the desired planewave in $\mathbb{D}_b$ is θ and is reproduced by loudspeakers positioned on an arc of angle ϕ_L and radius R_l with the first loudspeaker starting at angle ϕ.

In the orthogonal basis expansion approach to multizone soundfield reproduction [11], a soundfield function $S(\mathbf{x},k)$ that fulfils the wave equation, where $\mathbf{x} \in \mathbb{D}$ is an arbitrary spatial sampling point and k is the wavenumber of the soundfield, is written as a summation of an orthogonal set of the Helmholtz equation solutions [12] as

$$S(\mathbf{x},k) = \sum_n C_n(k,w) G_n(\mathbf{x},k) \tag{1}$$

where $\{G_n\}$ is a series of weighted basis functions and given a desired soundfield, $S^d(\mathbf{x},k)$, the expansion coefficients are derived as $C_n(k,w) = \int_{\mathbb{D}} S^d(\mathbf{x},k) G_n^*(\mathbf{x},k) w(\mathbf{x}) d\mathbf{x}$ using a weighted inner-product, where $w(\mathbf{x})$ represents the weighting function. The weighted QR factorisation may also use a weighted inner-product. Here, the weighting function, $w(\mathbf{x})$, can be written as $w(\mathbf{x}_b) = w_b$, $w(\mathbf{x}_q) = w_q$ and $w(\mathbf{x}_u) = w_u$, where $\mathbf{x}_b \in \mathbb{D}_b$, $\mathbf{x}_q \in \mathbb{D}_q$ and $\mathbf{x}_u \in \mathbb{D} \cap (\mathbb{D}_b \cup \mathbb{D}_q)'$, indicating different values for the weights within each of the bright ($\mathbb{D}_b$), quiet ($\mathbb{D}_q$), and unattended zones, respectively as illustrated in Figure 1.

Following a QR factorisation on a set of N planewaves, $F_n(\mathbf{x},k)$, arriving from angles $0 \to 2\pi$, (1) becomes,

$$S(\mathbf{x},k,w) = \sum_j P_j(k,w) F_j(\mathbf{x},k) \tag{2}$$

where the coefficients for the wavefields are $P_j(k,w) = \sum_n C_n(k,w)(\mathbf{R}^{-1})_{jn}$, $j \in \{1,\ldots,N\}$, $n \in \{1,\ldots,N\}$ and $\mathbf{R}$ is the upper triangular matrix from the QR factorisation [6].

The reproduction of the soundfield in (2) at a particular location can then be expressed as a summation of the discontinuously located loudspeaker signals [3],

$$S_{disc}^a(\mathbf{x},k,w) = \sum_{l=1}^{L} \omega_l(k,w) \frac{i}{4} H_0^{(1)}\left(k \left\| \mathbf{x}_l - \mathbf{x} \right\|\right) \tag{3}$$

where $\omega_l(k,w)$ is the l^{th} complex loudspeaker weight, L is the number of loudspeakers, $i = \sqrt{-1}$, $H_0^{(1)}(k\,\|\cdot\|)$ is a zeroth-order Hankel function of the first kind [12] and $\mathbf{x}_l$ is the position of the l^{th} loudspeaker. The complex loudspeaker weights are defined as,

$$\omega_l(k,w) = \sum_{m=-M}^{M} 2\left(i\pi H_m^{(1)}(kR_l)\right)^{-1} \left(\sum_j P_j(k,w) i^m e^{-im\phi_{pw}}\right) e^{im\phi_l} \Delta\phi_s \tag{4}$$

where $M = \lceil kR \rceil$ is the truncation length [3], R and R_l are from Figure 1, $\phi_{pw} = (j-1)\Delta\phi$ are the wavefield angles, $\Delta\phi = 2\pi / J$, ϕ_l is the angle of the l^{th} loudspeaker from the x-axis and $\Delta\phi_s$ is the angular spacing of the loudspeakers. The minimum number of loudspeakers to avoid aliasing is given by,

$$L \geq 2\left\lceil eRk_{max}/2 \right\rceil + 1 \tag{5}$$

where L is the minimum number of required loudspeakers, e is Euler's number and k_{max} is the wavenumber of the highest frequency (where frequency, $f = k \times c$ [12] and $c = 343 ms^{-1}$ is the speed of sound). See [3] for further details.

3. WEIGHTED MULTIZONE WIDEBAND SOUNDFIELDS

A wideband soundfield is described here as a linear combination of planewaves corresponding to each source frequency, similar to a Fourier series [12] and the approach of [8], [9]. The pressure generated at any point in the reproduced soundfield is given by,

$$\hat{p}(\mathbf{x},w) = \sum_k^K S_{disc}^a(\mathbf{x},k,w) \tag{6}$$

where there are K different sinusoidal components. Multiple nested summations are required to derive $\hat{p}(\mathbf{x},w)$ as:

$$\hat{\rho}(\mathbf{x},w) = \sum_k^K \sum_l^L \delta_{kl}(\mathbf{x}) \sum_{m=-M}^{M} \gamma_{km} \sum_j^J \beta_m \sum_n^N \alpha_{jn} \sum_{\mathbf{x}}^{\mathbb{D}} S^d(\mathbf{x},k) G_n^*(\mathbf{x},k) w(\mathbf{x}) \tag{7}$$

where $\alpha_{jn} = (\mathbf{R}^{-1})_{jn}$, $\beta_m = i^m e^{-im\phi_{pw}}$, $\gamma_{km} = 2e^{im\phi_l}\Delta\phi_s / (i\pi H_m^{(1)}(kR_l))$ and $\delta_{kl}(\mathbf{x}) = iH_0^{(1)}(k\,\|\,\mathbf{x}_l - \mathbf{x}\,\|)/4$. Here, a summation occurs for every sample in $\mathbb{D}$, for N planewaves, J wavefields, $2M+1$ modes, L loudspeakers and K sinusoidal components.

This paper extends the approach of [6] to allow for frequency-dependent weighting functions such that the leaked frequency spectrum can be controlled. This may benefit in cases where occlusion is a problem [9] and may improve the perceivable error compared to a single zone weight. To weight signals dynamically for each frequency, during the soundfield construction in (3), loudspeaker weights and soundfield samples are obtained per time-frequency component for a free field reproduction scenario (future work will look into reproduction in reverberant rooms). For example, if a signal from an arbitrary location in the reproduction area, defined in (3), was to be synthesised,

$$\hat{Y}_F(\mathbf{x},k) = \left| S_{disc}^a(\mathbf{x},k,w) \right| Y_F(k) \tag{8}$$

where $\hat{Y}_F(\mathbf{x},k)$ is the frequency domain signal at an arbitrary location, $\mathbf{x}$, in the reproduction region, $\mathbb{D}$, $Y_F(k)$ is obtained from the short-time Fourier transform of the windowed frame of input $y(n)$ and $|\cdot|$ denotes the absolute value. If loudspeaker signals are preferred to be synthesised then $|S_{disc}^a(\mathbf{x},k,w)|$ can be replaced with

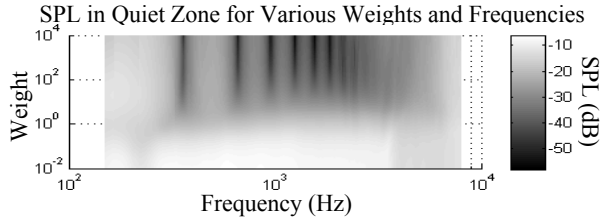

Figure 2 – Example high resolution table of values for SPL.

$\omega_l(k,w)$ from (4) and $\hat{Y}_F(\mathbf{x},k)$ becomes $\hat{Y}_{Fl}(k)$. Time domain loudspeaker signals, $\hat{y}_l(n)$, and/or soundfields, $\hat{y}(\mathbf{x},n)$, are derived from (8) via overlap-add reconstruction [13].

4. ORTHOGONAL PLANEWAVE SELECTION

The wavefield coefficients depend on the expansion coefficients and the upper triangular matrix resulting from QR factorisation computed on a set of N planewave functions arriving from angles $0 \rightarrow 2\pi$. Hence, the computation increases with both N and the spatial sampling density of the reproduction zone. A larger N results in a more accurate reproduction, however, in practice, this also results in poorly conditioned upper triangular matrices. A lower N causes spatial aliasing. The effect of N is also frequency dependent. Selecting N for the layout in Figure 1 is chosen to balance the impact of spatial aliasing and ill-conditioning:

$$N(f)=\left(N_c(f_{\max})-N_c(f_{\min})\right)\left(f_{\max}-f_{\min}\right)^{-1.2}\left(f-f_{\min}\right)^{1.2}+N_c(f_{\min}) \quad (9)$$

where f_{min} and f_{max} are the minimum and maximum frequencies, respectively, $N_c(f)$ is the largest number of planewaves that satisfies $\kappa(\mathbf{R}(f)) < 10^{10}$ or $\min(\epsilon_M^q(f))$ for a given frequency, f. Here, $\kappa(\mathbf{R})$ is the condition number of the upper triangular matrix, ϵ_M^q is the mean squared error (MSE) in the quiet zone and is limited to $-\infty < \epsilon_M^q < -60dB$. This work focuses on reproducing small room sized soundfields with a bandwidth $f \in 150Hz \rightarrow 8kHz$. Due to the size of the soundfields, zone weighting for $f < 150Hz$ has a negligible effect, as can be seen to trend in Figure 2. Analysis shows that $N_c(150) \approx 30$ and $N_c(8000) \approx 300$ based on (9). This allows the wideband system to be synthesised using the orthogonal basis expansion method with minimal error caused by the selection of N.

5. LOOK-UP TABLE BASED SYNTHESIS AND WEIGHTING

It is computationally demanding to construct a weighted multizone soundfield using the methods of Sections 2 and 3 due to the QR factorisation involved for all time-frequency components (e.g. a three second audio file sampled at $16kHz$ requires approximately 48×10^3 independent reproductions). In order to best make use of these repeated reproductions, the loudspeaker weights and soundfield pressure samples can be reproduced and stored for later use. Once enough values have been stored, interpolation between them can further reduce computation and error caused by quantised values. In this paper we propose the use of Look-Up Tables (LUT) to store pre-determined weighted soundfield values to be used for a given setup or wideband reproduction, with an example shown in Figure 2. A LUT may be described as a matrix of soundfield reproduction values for a given frequency and weight range,

$$\mathbf{A}_{uv} = \begin{bmatrix} S^a_{disc}\left(\mathbf{x},k_{min},w_{min}\right) & \dots & S^a_{disc}\left(\mathbf{x},k_{min},w_{max}\right) \\ \vdots & \ddots & \vdots \\ S^a_{disc}\left(\mathbf{x},k_{max},w_{min}\right) & \dots & S^a_{disc}\left(\mathbf{x},k_{max},w_{max}\right) \end{bmatrix} \quad (10)$$

where $S^a_{disc}(\mathbf{x},k,w)$ from (3) is a soundfield reproduction value for wavenumber, k, and weighting, w, at $\mathbf{x} \in \mathbb{D}$ and $\mathbf{A}_{uv}$ is a LUT with u number of frequencies and v number of weights in the range $\{k_{min},\dots,k_{max}\}$ and $\{w_{min},\dots,w_{max}\}$, respectively. The set of frequencies is logarithmically spaced as it closely resembles the spacing of the Bark scale [13] and the set of weights is logarithmically spaced as it provides larger control ranges in the decibel scale. The table can be built for loudspeaker signals by replacing $S^a_{disc}(\mathbf{x},k,w)$ with $\omega_l(k,w)$ from (4).

In order to evaluate the error and perceptual effects of quantising and interpolating soundfield values, a comparison is made between two LUTs (see Section 6). The MSE is evaluated as the difference between the interpolated values of lower and higher resolution LUT values as,

$$\varepsilon_{LUT} = \left\|\mathbf{A}_{u'v'}\right\|^{-1} \sum_{u'v'} \left(\tilde{\mathbf{A}}_{u'v'} - \mathbf{A}_{u'v'}\right)^2 \quad (11)$$

where ε_{LUT} is the MSE for the given interpolated LUT, $\tilde{\mathbf{A}}_{u'v'}$, relative to the highest resolution LUT, $\mathbf{A}_{u'v'}$, u' is the highest frequency resolution, v' is the highest weight resolution, u and v are the set of frequency and weight resolutions to evaluate, respectively. The interpolated LUT, $\tilde{\mathbf{A}}_{u'v'}$, is a matrix of size $u'v'$ obtained from interpolation of a smaller matrix, $\mathbf{A}_{uv}$. In this work bilinear interpolation is used.

6. RESULTS

This section describes simulations of reproduced soundfields using the above approach evaluated using MSE and PESQ measures.

6.1. Evaluation Setup

The multizone soundfield layout of Figure 1 is evaluated, where $r = 0.3m$, $r_z = 0.6m$, $R = 1m$, $R_l = 1.5m$, $\theta = \sin^{-1}(r/2r_z) \approx 14.5°$ and $\pi \approx 3.14159^c$. This setup is similar to [6] and θ is chosen such that an evanescent planewave with instant decay would interfere with half the quiet zone resulting in a large range of weighting control with a slight occlusion problem. Signals sampled at 16 kHz are converted to the time-frequency domain using a Hamming window (50% overlap) and Fast Fourier transform (FFT) of length 1024. The tables are analysed for a reproduction that meets and doesn't meet the aliasing criteria for the minimum number of reproduction loudspeakers given by (5). The LUTs are built for a reproduction where $L = 16$, $\phi = \theta + \pi/2$ and $\phi_L = \pi$, referred here on as the aliasing setup. They are also built where $L = 65$ and $\phi_L = 2\pi$, referred to as the non-aliasing setup. The aliasing setup aliases above $4kHz$ and the non-aliasing setup aliases above $8kHz$. The aliasing setup is included to show that visualising the soundfield values can provide other benefits.

The tables are built with the pressures for all $\mathbf{x} \in \mathbb{D}_b \cap \mathbb{D}_q$ and averaged across $\mathbb{D}_b$ and $\mathbb{D}_q$, from which the soundfield zones can be approximated. The zone weights are chosen as $w_b = 1$ and $w_u = 0.05$ following [6] and the variable weight is w_q. The effect of w_q on the input signal is evaluated using (8), which can also be reversed to find a soundfield level to obtain a desired output. This

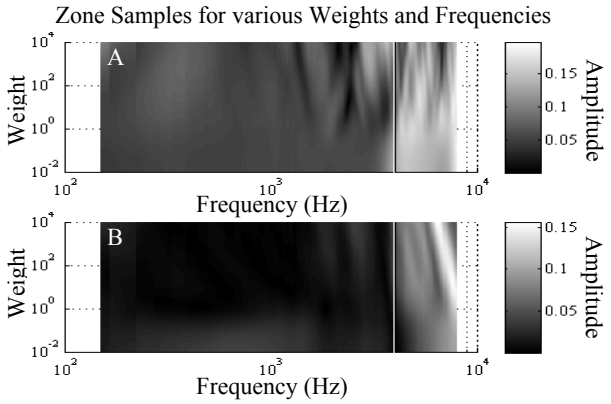

Figure 3 – LUT from the aliasing setup for the bright (A) and quiet (B) zones. The black and white line marks 4kHz aliasing.

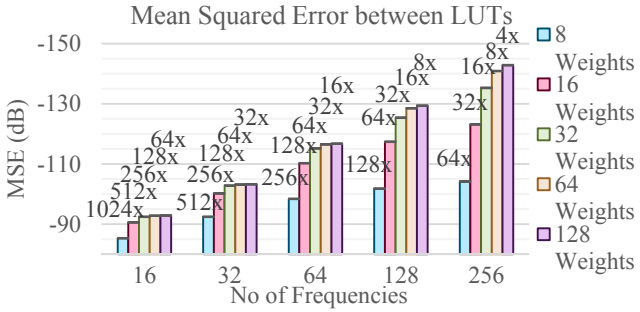

Figure 4 – MSE between different LUT resolutions. Labels show the relative complexity decrease from $\mathbf{A}_{u'v'}$.

soundfield level will map to a particular w_q in the LUT and, if implemented, to $\omega_l(k,w)$ that would reproduce the level.

Without interpolation or LUT's, the highest frequency resolution is 512 (based on the 1024 length FFT) and 256 different weight values (results in negligible reconstruction error). Each table is built for resolutions consecutively halving and decreasing in resolution down to 16 frequencies and 8 weights. In this work we have $w_q \in \{10^{-2},\ldots,10^{4}\}$ which extends the range used in [6]. The error between the different LUTs is evaluated using (11) where the highest resolution for frequency is $u' = 512$ and for weight $v' = u'/2$, and the set of frequency and weight resolutions to evaluate are $u = \{16,32,64,128,256\}$ and $v = u/2$, respectively.

The proposed approach is further evaluated using PESQ [10] to estimate the perceptual quality of the reproduced soundfields. Speech files for the evaluation are taken from the TIMIT corpus [14] where 20 files are chosen randomly. The male to female speaker ratio of these files is $50:50$. The reference signal for the PESQ algorithm is the original speech signal. PESQ values are obtained for the reproduced speech soundfields using the different resolution LUTs and then mapped to the PESQ Mean Opinion Score (MOS) [15]. These reproductions use $w_q = \{10^{-0.5},10^{0.5},10^{1.5},10^{2.5}\}$ such that they lie primarily in the centre of the interpolation regions. This allows the highest resolution LUT to be evaluated, however, due to the computational complexity is limited to four different weights.

6.2. Evaluation Results

Figure 3 illustrates the aliasing threshold described by (5). Here, it is expected that aliasing will occur above $4kHz$ and a

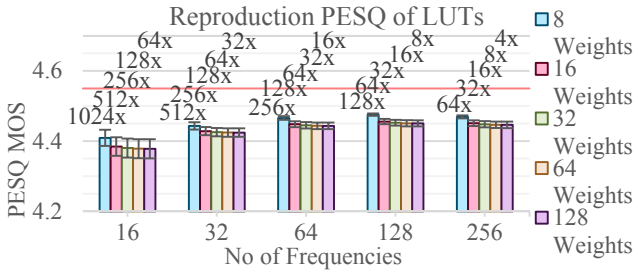

Figure 5 – PESQ MOS between weighted speech files reproduced by different LUTS with 95% confidence intervals. Labels show the relative complexity decrease from $\mathbf{A}_{u'v'}$. Red line indicates maximum mapped PESQ MOS.

pattern is noticeable, as noted by the black and white line, where it can be seen that the amplitude in both the bright and quiet zone becomes discernibly larger. It can also be seen that at about $8kHz$ in the quiet zone a significant aliasing occurs and when the weighting is increased it occurs at lower frequencies.

Analysing the MSE results between the different interpolation distances (Figure 4) indicate the lower resolution LUTs require significantly less computations than those of the higher resolution. This can be observed from the labels that show the relative decrease in the number of reproduced soundfields, which measured up to 1024 times less at just 0.10% the number of computations of the highest resolution LUT and with an MSE of $-85dB$, comparable to high end audio systems. In general, an increase in the interpolation distance increases the MSE.

As can be seen in Figure 5, the increased MSE caused by higher interpolation distances appears to have no discernible impact on the perceptual quality where the maximum mapped MOS is indicated by the red line. Figure 5 does show, however, a slight increase in the variability of the PESQ MOS, as indicated by the 95% confidence interval markers, where higher interpolation distances are required. This shows that interpolating the weighted soundfield values has an indiscernible perceptual effect on the reproduction and decreases the computational complexity of the problem with 1024 times less individual soundfield reproductions.

7. CONCLUSIONS

This paper proposed a method for building multizone soundfields for speech signals which allows dynamic control of the weighting between zones. We have proposed a method for reducing the computational effort involved when dynamically weighting zones for speech signals. We have also proposed a method for determining the number of planewaves required for multiple frequency systems utilising the orthogonal basis expansion approach. The method has been evaluated and shows indiscernible impact on perceptual quality of reproductions and decreased computational complexity. The evaluations show PESQ MOS lie around 4.4, MSE around $-85dB$ and 1024 times less individual reproductions. Also demonstrated in this paper is the use of LUTs for visualising speech soundfields and possible reproduction problems. For instance, media designers or engineers may benefit from the visualisation of these LUTs to better predict when aliasing or other phenomena occur in a system.

8. ACKNOWLEDGEMENTS

The authors would like to thank Wenyu Jin for his helpful insight into the orthogonal basis expansion method.

9. REFERENCES

[1] M. Poletti, "An Investigation of 2-D Multizone Surround Sound Systems," presented at the Audio Engineering Society Convention 125, 2008.

[2] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *JAES*, vol. 59, no. 10, pp. 721–734, Nov. 2011.

[3] Y. J. Wu and T. D. Abhayapala, "Theory and design of soundfield reproduction using continuous loudspeaker concept," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 107–116, 2009.

[4] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 93–96.

[5] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1711–1720, 2011.

[6] W. Jin, W. B. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," presented at the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013, pp. 311–315.

[7] H. Chen, T. D. Abhayapala, and W. Zhang, "Enhanced sound field reproduction within prioritized control region," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 2014, vol. 249, pp. 4055–4064.

[8] N. Radmanesh and I. S. Burnett, "Generation of isolated wideband sound fields using a combined two-stage lasso-ls algorithm," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 378–387, 2013.

[9] N. Radmanesh and I. S. Burnett, "Reproduction of independent narrowband soundfields in a multizone surround system and its extension to speech signal sources," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 461–464.

[10] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01)*, 2001, vol. 2, pp. 749–752 vol.2.

[11] G. H. Golub and C. F. van Van Loan, "Matrix computations (Johns Hopkins studies in mathematical sciences)," 1996.

[12] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. academic press, 1999.

[13] M. Bosi and R. E. Goldberg, *Introduction to digital audio coding and standards*. Springer, 2003.

[14] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon Technical Report N*, vol. 93, p. 27403, 1993.

[15] I. Rec, "P. 862.1: Mapping function for transforming P. 862 raw result scores to MOS-LQO," *International Telecommunication Union, Geneva*, vol. 24, 2003.