On the Extrapolation of Room Impulse Responses from Circular Measurements for Data Based Wave Field Synthesis

Achim Kuntz, Rudolf Rabenstein
Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Germany.
Email: {kuntz,rabe}@LNT.de

Introduction

The high quality realistic rendering of complex sound fields has always been one of the main motivations for the research on new sound reproduction systems. This contribution discusses the optimum preprocessing of room impulse responses for the spatial reproduction technique called wave field synthesis (WFS).

WFS is a reproduction technique that allows to control the sound field in a large listening area enclosed by a large number of loudspeakers [1] and is suitable for the reproduction of spatially complex sound fields. This paper is concerned with *data based WFS* systems, which use measured room impulse responses (RIRs) to auralize acoustic environments. The processing of measured RIRs for data based WFS is obvious in the case of an idealized measurement system that provides noise free data and an arbitrarily high spatial resolution [2]. For real world measurements, however, care has to be taken to account for the errors introduced by the RIR measurement. The processing of the impulse responses is topic of the following sections.

Room Impulse Response Processing

For data based WFS the loudspeaker driving signals are calculated as a multichannel convolution of dry recordings with RIRs form the positions of the sources in the acoustic scene to the loudspeaker positions [3]. Usually, the RIRs are not measured for the loudspeaker setup of a particular WFS system. Instead, the RIRs are measured with high spatial resolution and are then decomposed into a geometry independent field description. The RIRs can be extrapolated from this description.

To obtain the needed high spatial resolution, the RIRs have to be measured at a large number of positions. A very efficient approach is to mount a single microphone on a rotating stand and perform sequential measurements for a large number of angles. The RIRs measured on the circle can be interpreted as a signal in polar coordinates $P(r_0, \phi)$ and can thus be expanded as a Fourier series with coefficients $\mathring{P}_m(r_0)$. Following processing steps involve the decomposition of the RIR data into circular harmonics and the calculation of a plane wave decomposition of the impulse response data. The plane wave decomposition is the final format that allows to extrapolate the RIR data to the loudspeaker positions $\mathbf{x}_{\mathrm{LS}}$. This extrapolation step also incorporates the geometry of the WFS reproduction system. The measuring and processing scheme described above is depicted in Fig. 1.

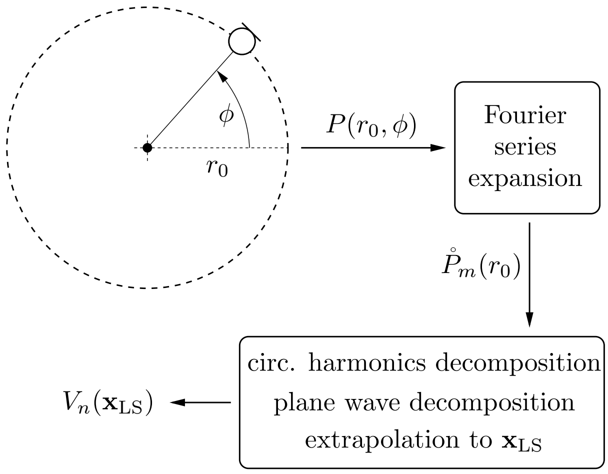

Figure 1: Measurement and extrapolation of room impulse responses for data based WFS systems.

Limitations due to Measurement Imperfections

Because of the microphone noise introduced in the measurement step and because of the finite number of microphone positions, S, only an approximation of the Fourier series coefficients can be captured,

$$\mathring{P}_{m,d} = \mathring{P}_m + \sum_{M \neq 0} \mathring{P}_{(m+MS)} + \mathring{N}. \qquad (1)$$

This equation shows that additionally to the wanted signal $\mathring{P}_m$ there are two error terms: The sum over M describes the modal repetitions due to the angular sampling, i. e. it describes modal aliasing. $\mathring{N}$ denotes the sensor noise in the modal domain (m, ω). We can assume spatially white noise so that $\mathring{N}$ does not depend on the mode number m.

For general wave fields it can be shown that the Fourier series coefficients show a modal low pass characteristics following the magnitude of the Bessel function, i. e. $|\mathring{P}_m| \approx |J_m(r_0 \cdot \omega/c)|$. Since J_m rapidly tends towards zero for $|m| > (r_0 \cdot \omega/c)$ and since the noise term is independent of m, the sensor noise $\mathring{N}$ leads to a very low SNR in the modal domain for modes $|m| > (r_0 \cdot \omega/c)$. Additionally, the aliasing terms cause large errors for high frequencies and high modal numbers m. Therefore, a preprocessing step has to take place before the RIR data can be further processed into a plane wave decomposition.

Common Approaches for Mode Selection

Usually, the preprocessing in the modal domain consists of a selection of usable Fourier series coefficients that show a sufficiently high SNR. To be on the safe side and avoid any noise or aliasing errors, it is of course possible to generously cut away all modes that could contain large errors. However, discarding modes implies discarding parts of the wanted signal and thus also causes errors. It is therefore desirable to preserve as much information of the RIR measurements as possible.

One commonly used possibility to discard the noisy modes is to limit the amplification in the calculation of the circular harmonics decomposition coefficients to a maximum value $A_{\max}$. Similar results can be obtained by discarding all modes for $|m| \gtrapprox (r_0 \cdot \omega/c)$. Yet, in both approaches an optimum selection criterion is unclear.

The modes containing aliasing can be selected using a hard decision: Either all modal components are kept including all aliasing errors, or all terms containing aliasing are discarded (e. g. using a hard modal limit according to $|m| < (S - r_o \cdot \omega/c - N_b)$ with appropriate N_b). Both approaches show distinct drawbacks: Keeping all modes results in strong spatial aliasing, which degrades the spatial accuracy of the WFS reproduction and also causes sound coloration. On the other hand, fully discarding a large number of modes limits the possible extrapolation positions to a smaller region [4]. Also, the selection of an optimum value for N_b is unclear.

Modal Optimal Filtering

As shown in the section before, the common approaches for the selection of usable modes require the determination of constants like $A_{\max}$ or N_b. Yet, the best possible values for these constants remain unknown. Thus, we propose to apply an optimal filter which minimizes the signal error in the modal domain,

$$H_{\text{opt}}(\omega, m) = \frac{\Phi_s(\omega, m)}{\Phi_s(\omega, m) + \Phi_e(\omega, m)}, \tag{2}$$

where Φ_s and Φ_e are the power spectral densities of the wanted signal $\mathring{P}_m$ and of the error terms. When considering the field of a single plane wave and disregarding higher order aliasing terms (the first order terms are strongest), it turns out that $H_{\text{opt}}(\omega, m)$ can be calculated independently of the RIR data. Only the SNR at the microphones and the array geometry have to be known. Application of the optimal filter yields the minimum error energy. The optimal filter allows to keep signal contributions in the aliased region where the SNR is positive, i. e. where the wanted signal dominates the aliasing terms.

Weighting factors can be applied to the error terms to provide a high flexibility of the optimal filter,

$$\Phi_e(\omega, m) = G_n \Phi_{\text{noise}}(\omega, m) + G_a \Phi_{\text{alias}}(\omega, m). \tag{3}$$

The minimum error power is obtained for $G_n = G_a = 1$. The weighting factor G_a can be used e. g. to account for different aliasing treatments of different WFS systems. Also frequency dependent factors may be profitable.

Results

Fig. 2 shows results for the extrapolation of a monofrequent plane wave for a simulated microphone array: $f = 4.7$kHz, $r_0 = 75$cm, $S = 100$ microphone positions, white noise resulting in SNR $= 50$dB at the microphones.

The left top sub-figure shows a simulated optimal extrapolation result which would be obtained in an aliasing free case. It serves as the reference field. The other three subfigures show the absolute difference to the reference field for different values of G_a. Using a low value $G_a = 0.05$ results in keeping all aliased modes. The extrapolation for $G_a = 1$ yields the minimum error energy. A high value $G_a = 1000$ results in discarding all aliased modes.

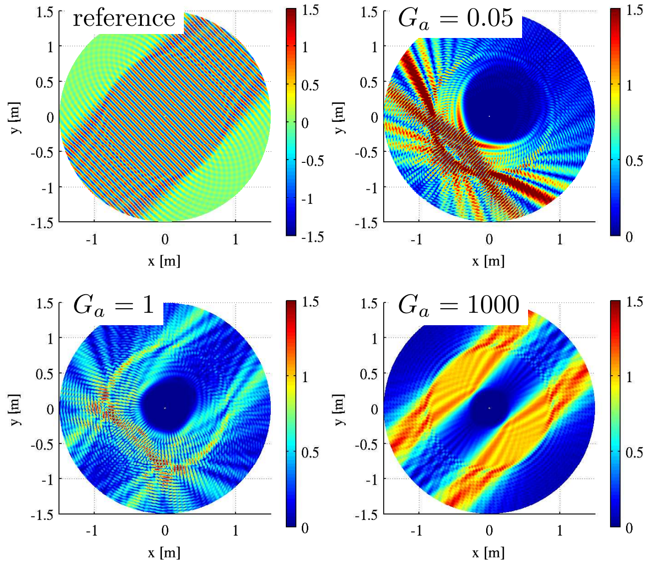

Figure 2: Extrapolation errors relative to the reference field as described in the text. Results are shown for different aliasing error weighting factors G_a. All signals are in linear scale.

Summing up, the variation of G_a allows to flexibly cross-fade between the two conventional aliasing treatment approaches. H_{opt} provides an optimal solution as well as it includes the conventional approaches.

References

[1] A.J. Berkhout, D. de Vries, and P. Vogel, Acoustic control by wave field synthesis, JASA, vol. 93, no. 5, pp 2764-2778, May 1993.

[2] E. Hulsebos, D. de Vries, and E. Bourdillat, Improved Microphone Array Configurations for Auralization of Sound Fields by Wave Field Synthesis, 110th AES Convention, Amsterdam, May 2001.

[3] S. Spors, H. Teutsch, A. Kuntz, R. Rabenstein, Sound Field Synthesis, in Audio Signal Processing for Next-Generation Multimedia Communication Systems, Y. Huang and J. Benesty (Eds), Kluwer Acad., 2004.

[4] A. Kuntz and R. Rabenstein, Limitations in the Extrapolation of Wave Fields from Circular Measurements, EUSIPCO'07, Poznan, Sep. 2007.