BINAURAL REPRODUCTION OVER LOUDSPEAKERS USING A MODIFIED TARGET RESPONSE

Ismael Nawfal and Joshua Atkins

Beats Electronics, LLC
1601 Cloverfield Blvd Ste 5000N
Santa Monica, CA 90404, USA

{ismael.nawfal, josh.atkins}@beatsbydre.com

ABSTRACT

Crosstalk cancellation (XTC) is a technique that can be used to play binaural content, typically meant for headphone playback, over two or more loudspeakers. Though effective at creating a binaural spatial sound field at the listening position, many XTC algorithms introduce spectral coloration, suffer from spatial robustness issues and create filters that are unrealizable in practice. Past approaches to dealing with this issue rely heavily on regularization. In this work we propose a new topology for loudspeaker binaural rendering (LBR) that performs better than conventional techniques without the need for regularization commonly associated with crosstalk cancellation based binaural renderers (XTC-BR). We then explore the use of a proposed LBR in the context of multiple output channels. A method is investigated to further optimize the filter design process by selecting an appropriate modeling delay and filter length using methods practiced in XTC filter design.

1. INTRODUCTION

In order to obtain a binaural effect with the use of two or more loudspeakers, it is necessary to control the amount of crosstalk from each speaker that arrives at an individual's ears. For this to occur, a set of filters can be designed to reduce or eliminate the crosstalk at the listening position. This concept of using acoustic crosstalk cancellation (XTC) to achieve a binaural or spatial effect has been extensively explored since the original system was proposed by Atal and Schroeder in 1966 [1]. The concepts related to XTC can be extended to create virtual audio sources in space. The effectiveness of this type of crosstalk cancellation based binaural renderer (XTC-BR) is dependent on the filter design method employed, such as choice of regularization, cost function, as well as the physical constraints determined by the number of loudspeakers used and their geometry in space.

Loudspeaker span influences the effectiveness and robustness of a XTC or XTC-BR and has a large effect on the feasibility of different XTC filter design methods reported in the past [2, 3, 4, 5, 6]. Acoustically, closely angled loudspeakers tend to provide more robustness to head movement, but compromise the achievability of XTC at lower frequencies [5, 6] and typically require a considerable amount of regularization. By contrast, widely angled loudspeakers tend to reproduce lower frequencies at the expense of spatial robustness. Using more than two drivers allows for multiple spans which improves the effectiveness and robustness of a crosstalk canceler across a wider frequency range [7]. Similarly, the distance between drivers in a XTC system plays a role in determining the upper frequency limit at which XTC is achievable [4]. To address this, a linear array was suggested in [4] using either a multi-band approach processing band-passed content to appropriately spaced loudspeakers or utilizing a constant speaker spacing, $\mathbf{d_s}$ in Figure 1, corresponding to the highest frequency of interest. A similar relationship between speaker distance and reproducible frequency was examined further in [8], where a conceptual transducer is proposed whose position varies with frequency. In [9], a uniformly spaced linear array was introduced that implemented XTC utilizing a sub-band regularization scheme. A uniformly spaced linear array is illustrated in Figure 1. Past approaches to XTC filter design include the use of Tikhonov regularization [10], using plane-wave approximations of the HRTF [11], jointly optimizing for a set of measurement positions [12], using sub-band approaches to achieve cancellation only in a certain frequency range [9], and adjustment of the error-norm to better match perception [13]. The signal processing associated with a multichannel XTC

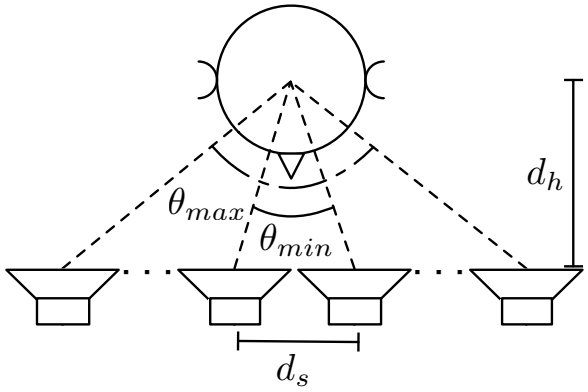

Figure 1: An example of a typical listening position with an M speaker uniform linear speaker array consisting of a minimum span θ_{min}, maximum span θ_{max}, listener distance d_h, and speaker spacing d_s.

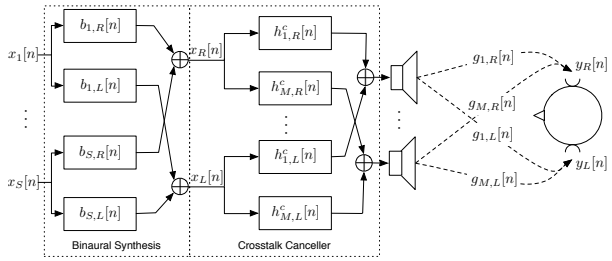

Figure 2: A typical XTC-BR system for simulating binaural sources over loudspeakers with S sources and M loudspeakers. The filters $h^c[n]$ represent XTC filters while filters $b[n]$ represent the HRTFs of the desired angle to be rendered.

is illustrated in Figure 2.

Many alternative methods to XTC have been proposed in the literature for sound field synthesis over two or more loudspeakers with most approaches focusing on large loudspeaker arrays. The most widely used methods are Higher-Order Ambisionics (HOA), Vector Based Amplitude Panning (VBAP), and Wave Field Synthesis (WFS) [14, 15, 16]. HOA and WFS are often referred to as analytical methods since they formulate and solve the synthesis problem exactly using approaches from physical acoustics. Of these three, only WFS has been studied extensively for reproduction over linear loudspeaker arrays [17, 18] (HOA and VBAP focus on circular and spherical arrays). More general methods that treat sound field reproduction as a numerical inverse problem have also been studied extensively in recent work [19, 20, 21].

The approach presented in this paper shares more commonality with the numerical approaches than the analytical methods like HOA and WFS, but draws inspiration from the XTC literature where the goal is focused on binaural reproduction. Here we study the effectiveness of a proposed loudspeaker binaural renderer (LBR), using a novel method illustrated in Figure 3, to simulate virtual sources at arbitrary angles. In the conventional approach, the filters for a XTC-BR are designed with a target flat frequency response at the ipsilateral ear and a fully attenuated response at the contralateral ear and then convolved with HRTFs associated with a desired angle. While this works well at playing back an unknown binaural recording, it is suboptimal for the binaural simulation case, especially when regularization methods are employed. To address this, we propose a modified target response for the filter design whereby a set of LBR filters is designed directly. This approach has two motivations. Firstly, it leads to achievable binaural reproduction filters without the need for regularization at low-frequencies. Secondly, the necessary amount of crosstalk attenuation is over-estimated in the traditional scheme, as the binaural effect is typically achievable with 15-20 dB of channel separation depending on the source content used [22]. Through simulations we will show how this approach leads to better timbral reproduction without the need for traditional forms of regularization. We will also expand our work to the multichannel case shown in Figure 1. Additionally we will propose an optimization scheme for filter generation that ensures the faithful spatial reproduction of multichannel content while also reducing processing requirements. We will focus on the effect of modeling delay and filter length, parameters commonly explored in crosstalk filter design, on the filter design of our proposed LBR.

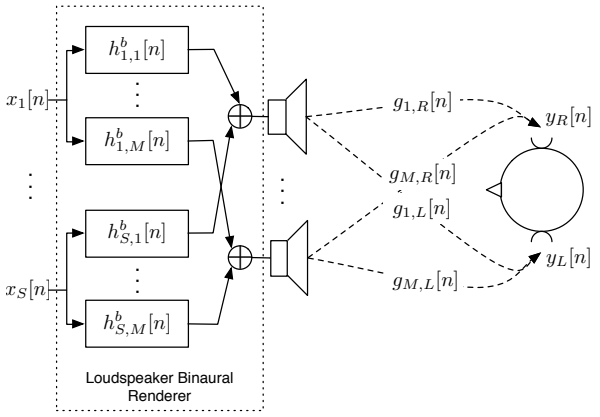

Figure 3: The proposed LBR system for simulating binaural sources over loudspeakers with S sources and M loudspeakers. The filter $h^b[n]$ represents the binaural rendering filter used to simulate a virtual source.

2. BACKGROUND AND THEORY

A conventional XTC system is shown in Figure 2. We define $x_s[n]$ as a monaural source signal, $x_L[n]$ and $x_R[n]$ as inputs to the crosstalk canceler, h^c_m as the crosstalk cancellation filter for loudspeaker m, $g_{m,L}$ and $g_{m,R}$ as the acoustic transfer functions between each loudspeaker and each ear, and $y_L[n]$ and $y_R[n]$ as the reproduced signals at each ear. From this, we can write the reproduced sound field at the listening position as

$$\begin{aligned} y_L[n] &= \sum_{m=1}^{M} g_{m,L} * (h^c_{m,L} * x_L[n] + h^c_{m,R} * x_R[n]) \\ y_R[n] &= \sum_{m=1}^{M} g_{m,R} * (h^c_{m,L} * x_L[n] + h^c_{m,R} * x_R[n]). \end{aligned} \tag{1}$$

where $*$ represents the convolution operator. The acoustic path models, $g_{m,L}$ and $g_{m,R}$, are typically chosen using the measured HRTFs for each path, but in some methods a plane-wave approximation is used as a simplification [11]. In this case, the ipsilateral propagation is simulated as $\delta[0]$ and the contralateral as $\alpha\delta[n-\tau_c]$, where $\delta[\cdot]$ is the delta function, α is a modeled attenuation factor, and τ_c is a modeled propagation delay.

The desired sound field at the listening position will be $d_L[n] = t_L * x[n]$ and $d_R[n] = t_R * x[n]$. Conventionally, the target reproduction filters, t_L and t_R, are chosen as the filters 0 and $\delta[n - D]$, where D is a modeling delay. This choice of filters aims at achieving zero signal at the contralateral ear and a delayed version of the signal at the ipsilateral ear, a scenario much like the headphone listening case. The modeling delay is important to ensure causality of the designed filters in the conventional design framework [12].

The filter design then aims to find the filters h_m such that a suitable choice of error function, $e(y_L[n], d_L[n], y_R[n], d_R[n])$, is minimized. Typically the ℓ_2-norm cost function is chosen. We focus the analysis in this work on the specific scenario where the crosstalk canceler is used with a binaural renderer (XTC-BR) to simulate a sound source from a given direction.

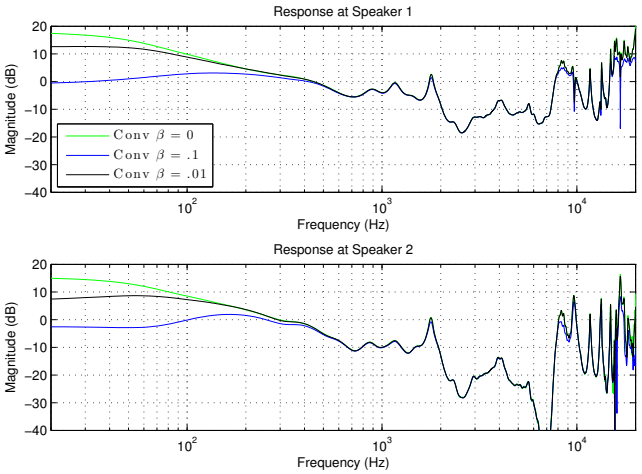

Figure 4: A comparison of XTC filter responses at the speaker generated using a conventional time domain method with regularization parameter $\beta = 0$, 0.1 and 0.01 for stereo reproduction using loudspeakers at a 5° angular span.

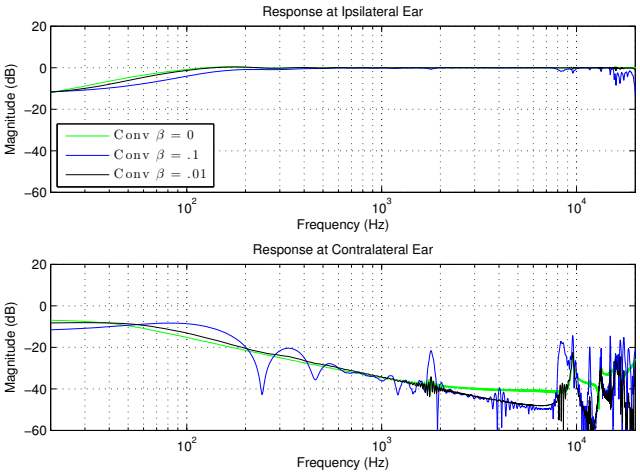

Figure 5: A comparison of XTC filter responses at the ear generated using a conventional time domain method with regularization parameter $\beta = 0$, 0.1 and 0.01 for stereo reproduction using loudspeakers at a 5° angular span.

2.1. Traditional Filter Generation Methodology

XTC filters are commonly calculated in the time domain by solving the following minimization, which solves for the filters $h^c_{I,1} \dots h^c_{I,M}$ and $h^c_{C,1} \dots h^c_{C,M}$,

$$\underset{\mathbf{h}^c}{\text{minimize}} \quad \|\mathbf{G}^c\mathbf{h}^c - \mathbf{t}^c\|_2 + \beta\|\mathbf{h}^c\|_2 \tag{2}$$

$$\mathbf{t}_I = [\underbrace{0, \dots, 0}_{D}, 1, \underbrace{0, \dots, 0}_{N_h-1}]$$

$$\mathbf{t}_C = [\underbrace{0, \dots, 0}_{D}, 0, \underbrace{0, \dots, 0}_{N_h-1}]$$

$$\mathbf{t}^c = [t^c_I, t^c_C]^T$$

$$\mathbf{g}^c_{m,I} = [g^c_{m,I}[0], \dots, g^c_{m,I}[N_g - 1], \underbrace{0, \dots, 0}_{D}]^T$$

$$\mathbf{g}^c_{m,C} = [g^c_{m,C}[0], \dots, g^c_{m,C}[N_g - 1], \underbrace{0, \dots, 0}_{D}]^T$$

$$\mathbf{G}^c = \begin{bmatrix} \text{convmtx}(\mathbf{g}^c_{1,I}) & \dots & \text{convmtx}(\mathbf{g}^c_{M,I}) \\ \text{convmtx}(\mathbf{g}^c_{1,C}) & \dots & \text{convmtx}(\mathbf{g}^c_{M,C}) \end{bmatrix}$$

$$\mathbf{h}^c_m = [h_m[0], \dots, h_m[N_h - 1]]$$

$$\mathbf{h}^c = \begin{bmatrix} \mathbf{h}_1 & \dots & \mathbf{h}_M \end{bmatrix}^T$$

where $\mathbf{t}_I$ and $\mathbf{t}_C$ are the desired responses at the ipsilateral and contralateral ears respectively. Similarly $\mathbf{g}^c_{m,I}$ and $\mathbf{g}^c_{m,C}$ are the actual HRTF responses at the ipsilateral and contralateral ears respectively. The notation $\|\cdot\|_2$ is the ℓ_2-norm and convmtx$(\cdot)$ creates the acyclic convolution matrix of size $N_g + N_h - 1$ by N_h for a given vector. Variables N_g, N_h, and N_t correspond to the lengths of the acoustic path, filter, and desired response, respectively, and D represents the modeling delay that is used to ensure causality. The variable M refers to the number of loudspeakers.

The vector $\mathbf{t}^c$ consists of the perfect crosstalk cancelled response at the listener's ears. Tikhonov regularization is often used to avoid issues with matrix inversion where the parameter β controls the amount of regularization [11, 23] and is commonly used in a frequency dependent fashion [9, 24, 25]. Modifications to this form can be seen in [26], where a non-constant regularization parameter is calculated. The designed filters, $\mathbf{h}^c$, can be convolved with HRTF measurements from a particular source angle to simulate a binaural sound from a desired direction as done in XTC-BR and shown in Figure 2.

Alternatively, XTC filters can be designed in the frequency domain to allow for constraints that better match perception. Note that due to Parseval's theorem, the cost function is the same in time and frequency when using the ℓ_2-norm. These methods along with the typical regularization approaches have been explored in [9, 11, 27]. The problem in the frequency domain can be written as

$$\underset{\mathbf{h}^c}{\text{minimize}} \quad \|\underline{\mathbf{G}}^c\mathbf{F}_B\mathbf{h}^c - \underline{\mathbf{t}}^c\|_2 + \beta\|\mathbf{F}_B\mathbf{h}^c\|_2. \tag{3}$$

$$\mathbf{F}_B = \text{blkdiag}(\mathbf{F}, M)$$

$$\underline{\mathbf{t}}^c = [(\mathbf{F}\mathbf{t}_I)^T, (\mathbf{F}\mathbf{t}_C)^T]^T$$

$$\underline{\mathbf{G}}^c = \begin{bmatrix} \text{diag}(\mathbf{F}\mathbf{g}^c_{1,I}) & \dots & \text{diag}(\mathbf{F}\mathbf{g}^c_{M,I}) \\ \text{diag}(\mathbf{F}\mathbf{g}^c_{1,C}) & \dots & \text{diag}(\mathbf{F}\mathbf{g}^c_{M,C}) \end{bmatrix}$$

where $\mathbf{F}$ represents the discrete Fourier transform matrix of size $\max(N_g + N_h - 1, N_t)$ by N_h, diag$(\cdot)$ is the diagonal matrix formed from a vector, and blkdiag$(\cdot, M)$ is the block diagonal matrix formed by repeating a matrix along its diagonal M times. The underline notation represents a quantity in the frequency domain.

2.2. Proposed Filter Generation Method

We propose altering the conventional XTC-BR approach to use a modified target response for the filter design so that LBR filters are designed directly. This proposed method will generate well-behaved filters that do not require any constraints for regularization. Our design paradigm is illustrated further in Figure 3. Our proposed method, which solves for the filters $h^b_{L,1} \dots h^b_{L,M}$ and $h^b_{R,1} \dots h^b_{R,M}$, can be described as

$$\underset{\mathbf{h}^b}{\text{minimize}} \quad \|\mathbf{G}^b\mathbf{h}^b - \mathbf{t}^b\|_2 \tag{4}$$

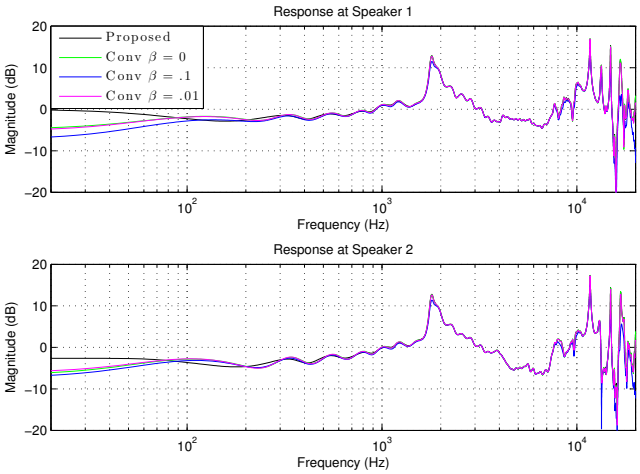

Figure 6: A comparison of the speaker responses of the LBR and XTC-BR methods with regularization parameter β= 0, 0.1 and 0.01, attempting to use loudspeakers at a 5° angular span to simulate a source at +/- 30° .

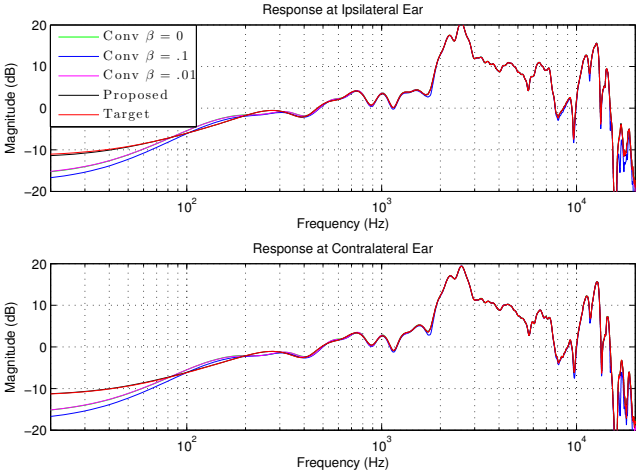

Figure 7: A comparison of the ear responses of the LBR and XTC-BR methods with regularization parameter β= 0, 0.1 and 0.01, attempting to use loudspeakers at a 5° angular span to simulate a source at +/- 30° .

$$\mathbf{t}_L^b = [\underbrace{0,\ldots,0}_{D}, t_L^b[0],\ldots,t_L^b[N_t-1], \underbrace{0,\ldots,0}_{N_{h-1}}]$$

$$\mathbf{t}_R^b = [\underbrace{0,\ldots,0}_{D}, t_R^b[0],\ldots,t_R^b[N_t-1], \underbrace{0,\ldots,0}_{N_{h-1}}]$$

$$\mathbf{t}^b = [t_L^b, t_R^b]^T$$

$$\mathbf{g}_{m,L}^b = [g_{m,L}^b[0],\ldots,g_{m,L}^b[N_g-1], \underbrace{0,\ldots,0}_{D}]^T$$

$$\mathbf{g}_{m,R}^b = [g_{m,R}^b[0],\ldots,g_{m,R}^b[N_g-1], \underbrace{0,\ldots,0}_{D}]^T$$

$$\mathbf{G}^b = \begin{bmatrix} \text{convmtx}(\mathbf{g}_{1,L}^b) & \ldots & \text{convmtx}(\mathbf{g}_{M,L}^b) \\ \text{convmtx}(\mathbf{g}_{1,R}^b) & \ldots & \text{convmtx}(\mathbf{g}_{M,R}^b) \end{bmatrix}$$

$$\mathbf{h}_m^b = [h_m[0],\ldots,h_m[N_h-1]]$$

$$\mathbf{h}^b = \begin{bmatrix} \mathbf{h}_1^b & \ldots & \mathbf{h}_M^b \end{bmatrix}^T$$

where $\mathbf{t}_L^b$ and $\mathbf{t}_R^b$ are the desired responses at the left and right ears respectively. Similarly $\mathbf{g}_{m,L}^b$ and $\mathbf{g}_{m,R}^b$ are the actual HRTF responses at the left and right ears respectively. The elements in $\mathbf{h}^b$ are the desired LBR filters that simulate a virtual source at a specific angle and $\mathbf{t}^b$ consists of the desired responses at the listener's ears. Note here we do not solve for XTC filters, but rather filters that best render a binaural source at a given angle. By solving directly, the filter design process is simplified and there is no need to apply both XTC and binaural rendering filters as is done in XTC-BR. Filters generated using this method were found to perform well when designed with a filter length correlating to the length of the source HRTF used. The generated LBR filters can be used to render virtual sources in space by calculating filter coefficients in advance with a desired spatial resolution and using interpolation techniques commonly used in XTC and XTC-BR [2].

3. SIMULATIONS

In this section we compare the XTC-BR method against the proposed LBR method. We study the application of Tikhonov regularization to the XTC and XTC-BR filter design problems and evaluate the simulated reproduction at each ear and the resulting filters in each case.

In this experiment we generate filters for loudspeakers measured with a span of 5° to simulate identical loudspeakers with a span of 60° using the conventional and proposed filter design methods. A loudspeaker span of 60° is considered standard for stereo reproduction of music. To test the efficacy of each method in widening a stereo image from 5° to 60° we study the ability of the designed filters to simulate virtual loudspeakers at +/- 30° . All filters were designed to be 1024 taps incorporating a modeling delay of 256 samples and were generated using time-domain design methods shown in Equation 2 and Equation 4. The HRTFs used were measured in an anechoic chamber known to have a cutoff of 40 Hz using a KEMAR manikin placed 2 m from a studio monitor with a near flat frequency response from 50 Hz to 20 kHz.

Firstly we examine the effect of the regularization parameter, β, on conventional XTC filter design. Figure 4 shows the conventional XTC filter generation utilizing different β values. As can be seen in the graph, the use of regularization can influence the frequency response of the designed filter. This can lead to a compromise in the robustness of the designed filter especially at low frequencies. Comparing the simulated response at the ear in Figure 5, it is clear that the use of a regularizer has an effect on perceived cross-path attenuation. An increase in β simultaneously reduces the power required by a speaker for playback and the level of cross-path attenuation.

The conventional XTC-BR method shown in Figure 2 utilizing different β values was compared against our proposed LBR method shown in Figure 3. Both methods attempt to simulate a source at +/- 30° utilizing loudspeakers with a 5° span. The simulated speaker and ear responses for both methods can be seen in Figure 6 and Figure 7 respectively. Both figures show that each method tends to match the desired target response well at mid and high frequencies. However, the proposed LBR method tends to match better below 200 Hz. Informal listening confirmed that the LBR method performed better than the conventional XTC-BR approach by better simulating a target angle with limited timbral arti-

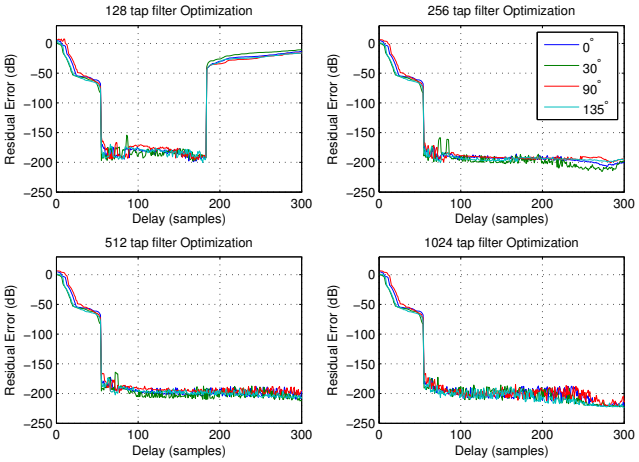

Figure 8: A comparison of the residual ℓ_2 errors for LBR filters at the left ear for 128, 256, 512 and 1024 taps rendering at 0° , 30° , 90° and 135° at different modeling delays. All filter lengths and target HRTF combinations suggest similar optimal modeling delays.

facts, specifically in the lower frequency range. This better performance was achieved without the use of any type of regularization.

4. APPLICATION

In the previous section, a comparison of the proposed LBR and conventional XTC-BR method of binaural reproduction of a single source was evaluated via simulation and informal listening. It was found that although both methods tended to match the binaural target well, the LBR method performed better particularly at lower frequencies. Having compared LBR and XTC-BR methods, in this section we continue our evaluation of our proposed LBR method by applying it to the context of an eight speaker uniform linear speaker array simulating multiple sources. Two parameters play a significant role in the filter optimization of the proposed LBR: modeling delay, D, and modeled filter length, N_h, as seen in Equation 4.

The HRTF measurements used in this section were taken using a KEMAR mannequin placed 2 m in front of an 8 speaker linear array in an anechoic chamber. Each speaker in the linear array consisted of a 10 cm woofer paired with a 2.5 cm tweeter crossing over at 2 kHz. The distance between drivers, $\mathbf{d_s}$, was 10 cm, which corresponds to θ_{min} = 3° and θ_{max} = 20° . Impulse responses were acquired using the logarithmic sine sweep method [28]. The target HRTFs were measured with the same studio monitor described in Section 3 at angles corresponding to those commonly associated with 7.1 surround sound playback: 0° , 30° , 90° and 135° . Measured responses were truncated to 200 samples, removing minor reflections that were observed within measurements and content residing near the noise floor.

4.1. Delay Optimization

A modeling delay is important to ensure causality of the designed filters in the XTC filter design [12]. It is stated in [12] that any causal modeling delay is usually considered acceptable and that a delay that falls in the middle of the causal region would be reasonable. An optimal modeling delay will minimize the least squares cost function while ensuring causality and robustness of the final filter design. The modeling delay is described as D and is appended to the beginning of the target HRTF response, $\mathbf{t}_s$, in Equation 4. In order to determine an optimal delay, the residual ℓ_2 error was calculated for various modeling delays for filter lengths of 128, 256, 512 and 1024 samples. As can be seen in Figure 8, the modeling delay remains relatively consistent for all filter lengths. It would be reasonable to assume, as in [12], that any modeling delay in the causal region would be acceptable, thus we choose the smallest modeling delay in the causal region in our study of filter length optimization which is 55 samples. Note that the modeling delay used is only relevant during the filter design process and does not impact the computational complexity of the resulting filters.

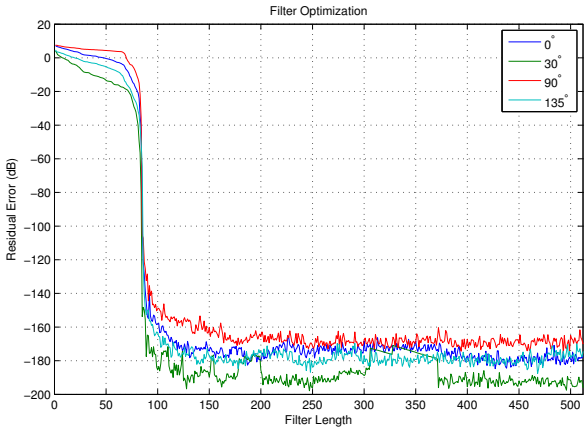

Figure 9: Residual ℓ_2 error for LBR filters at the left ear as a function of filter length using a modeling delay of 55 samples determined in Section 4.1.

4.2. Filter Length Optimization

Since the filter length used directly influences the computational complexity of real-time applications, shorter filter lengths are preferred. Using the conclusions in Section 4.1, a modeling delay of 55 samples is used in order to establish an optimal filter length. We will define the optimal filter length to be the filter length at which a significant reduction in residual ℓ_2 error occurs. Residual ℓ_2 errors associated with filter lengths were iteratively calculated and are shown in Figure 9. As expected, shorter filters tended to generate more residual ℓ_2 error, however there is a point at which the filter length no longer becomes a limiting factor. In the figure we can see the optimal filter length for our example is 86 taps, which corresponds to the point at which the residual ℓ_2 error reaches a level below -120 dB.

4.3. Joint Optimization

Figure 10 shows a three-dimensional representation of the optimization scheme described in Section 4. From Section 4.1 and Section 4.2 we suggest a modeling delay of 55 samples and a filter length of 86 taps for this particular problem. The assumption is that all areas that are shown as dark blue in Figure 10 would be appropriate selections for modeling delay and filter length combinations in order to generate binaural rendering filters due to their very low residual ℓ_2 errors.

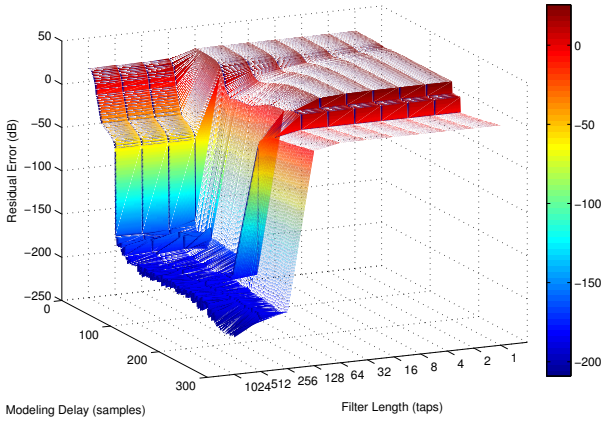

Figure 10: A waterfall plot showing the relationship between modeling delay, filter length and residual ℓ_2 error for generated LBR filters.

5. DISCUSSION

In order to confirm the selection of our optimal values of D = 55 samples and N_h = 86 taps, informal listening was conducted with the derived optimal modeling delay and filter length values. Though numerically shown to have minimum residual ℓ_2 error, perceptually these values did not provide the expected spatial rendering. A listening test was set up in order to determine the minimum modeling delay required in order to achieve the expected spatial rendering, ignoring residual ℓ_2 error data and the optimal modeling delay that was calculated previously. It was found that a modeling delay of approximately 100 samples, which fell well within the causal region, was the minimum necessary to achieve the correct spatial rendering of all angles (0° , 30° , 90° and 135°) with 1024 tap binaural filters.

The derived optimal filter length of 86 taps also failed to provide correct rendering, and thus the filter length was reduced incrementally from 1024 taps in order to determine a minimum length that would provide the expected spatial rendering. A filter length of 200 samples was subjectively determined to provide the correct spatial rendering with no significant timbral artifacts utilizing a modeling delay of 100 samples. Interestingly, this value corresponds to the lengths of the measured and target HRTF responses, N_g and N_t respectively (from Equation 4), which were used during the filter design process. This indicates that the length of the HRTF measurements used in the design process may play a more dominant role in determining the optimal length of a binaural rendering filter than residual ℓ_2 error.

Despite the residual ℓ_2 error data that was derived, subjective listening indicates that residual ℓ_2 error is not a sufficient metric in determining the effectiveness of filters to be used for binaural rendering. Though values of modeling delay and filter length have low residual ℓ_2 error and fall in the causal region they did not necessarily generate perceptually valid filters. This is in contrast to [12], which suggested selecting a modeling delay in the middle of the causal region or a modeling delay with the lowest residual ℓ_2 error as a rule of thumb. These suggestions made in [12] were not validated with listening tests as done here. As an example, note the modeling delay versus residual ℓ_2 error for a 128 tap filter in Figure 8 where we can see a clear causal modeling delay region. Despite this, we found that no causal modeling delay was capable of providing a perceptually acceptable 128 tap filter. Due to these conclusions, we propose that residual ℓ_2 error should be used as baseline in conjunction with perceptual evaluation in order to ensure that all spatial rendering is intact after optimization.

6. CONCLUSIONS

In this paper we have shown a technique for reproducing a binaural sound source with minimal artifacts. In departure from previous methods with crosstalk cancelers, the proposed LBR approach attempts to directly design the optimal reproduction filter for a given source angle. This reduces the need for regularization and simplifies the overall filter design process. A comparison of the proposed LBR and conventional XTC-BR reproduction of a single source was evaluated via simulation and informal listening evaluation. It was found that although both methods tended to match the binaural target well, the LBR method performed better at lower frequencies.

We have also applied the LBR method to the context of reproducing a binaural sound source using multiple drivers with minimal artifacts. The system was implemented and validated with an 8-channel uniform linear array. We have shown that both the choice of delay applied to the target response and desired filter length play a significant role in the optimization of the residual ℓ_2 error function. Despite determining a method for minimizing residual ℓ_2 error while reducing modeling delay and filter length, informal listening evaluations have shown that residual ℓ_2 error may not be the only factor that should be considered when attempting to optimize binaural rendering filters. Instead, modeling delay and filter length should be validated perceptually, taking filter design residual ℓ_2 error as a tool for narrowing down the selection process.

7. REFERENCES

[1] B.S. Atal and M.R. Schroeder, “Apparent sound source translator,” *US Patent 3,236,949*, Feb. 1966.

[2] W.G. Gardner, *3-D Audio Using Loudspeakers*, Ph.D. thesis, Massachusettes Institute of Technology, Sept. 1997.

[3] J.M. Jot, V. Larcher and O. Warusfel, ”Digital Signal Processing Issues in the Context of Binaural and Transaural Stereophony,” *Audio Engineering Society Convention 116*, February 1995.

[4] D.B. Ward and G.W. Elko, “Optimum loudspeaker spacing for robust crosstalk cancellation,” *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 6, pp. 3541–3544, May 1998.

[5] D.B. Ward and G.W. Elko, “Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation,” *IEEE Signal Processing Letters*, vol. 6, no.5, pp. 106–108, May 1999.

[6] O. Kirkeby, P.A. Nelson, and H. Hamada, “The stereo dipole - A virtual source imaging system using two closely spaced loudspeakers,” *Journal of the Audio Engineering Society*, vol. 46, pp. 387–395, May 1998.

[7] J. Bauck and D. Cooper ”Generalized Transaural Stereo and Applications,” *Journal of the Audio Engineering Society*, vol. 44, No.6, Sep. 1996.

[8] T.Takeuchi and P.A. Nelson, "Optimal source distribution for binaural synthesis over loudspeakers," *Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 2786–2797, Dec. 2002.

[9] M.R. Bai and C.C. Lee, "Development and implementation of crosstalk cancellation system in spatial audio reproduction based on subband filtering," *Journal of Sound and Vibration*, vol. 290, no. 3–5, pp.1269–1289, Mar. 2006.

[10] O. Kirkeby, P.A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, Mar. 1998.

[11] O. Kirkeby, P.A. Nelson, and H. Hamada, "Local sound field reproduction using two closely spaced loudspeakers," *Journal of the Acoustical Society of America*, vol. 104, no. 4, pp. 1973–1981, Oct. 1998.

[12] D.B. Ward, "Joint least squares optimization for robust acoustic crosstalk cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, pp. 211–215, Mar. 2000.

[13] H. Rao, V.J. Mathews, Y.-C. Park, "A Minimax Approach for the Joint Design of Acoustic Crosstalk Cancellation Filters," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 2287–2298, Nov. 2007.

[14] A.J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *Journal of the Acoustical Society of America*, vol. 93, pp. 2764, May 1993.

[15] J. Daniel and S. Moreau, "Further Study of Sound Field Coding with Higher Order Ambisonics," *Audio Engineering Society Convention 116*, May 2004.

[16] V. Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, June 1997.

[17] S. Spors, "Spatial Aliasing Artifacts Produced by Linear Loudspeaker Arrays used for Wave Field Synthesis," *Second IEEE-EURASIP International Symposium on Control, Communications, and Signal Processing*, pp. 1–4, Nov. 2006.

[18] S. Spors and J. Ahrens, "Analysis and Improvement of Pre-Equalization in 2.5-Dimensional Wave Field Synthesis," *Audio Engineering Society Convention 128*, May 2010.

[19] F.M. Fazi and P.A. Nelson, "Sound field reproduction as an equivalent acoustical scattering problem.," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3721–3729, Nov. 2013.

[20] G. Lilis, D. Angelosante, and G. Giannakis, "Sound Field Reproduction using the Lasso," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 1902–1912, Nov. 2010.

[21] M. Kolundzija, C. Faller, and M. Vetterli, "Reproducing Sound Fields Using MIMO Acoustic Channel Inversion," *Journal of the Audio Engineering Society*, vol. 59, no. 10, pp. 721–734, Oct. 2011.

[22] Y.L. Parodi and P. Rubak, "A Subjective Evaluation of the Minimum Audible Channel Separation in Binaural Reproduction Systems Through Loudspeakers," *Audio Engineering Society Convention*, May 2010.

[23] M.R. Bai, C.T. Tung, and C.C. Lee, "Optimal Design of Loudspeaker Arrays for Robust Cross-talk Cancellation Using the Taguchi method and the genetic algorithm," *Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 2802-2813, May 2005.

[24] O. Kirkeby and P.A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no.7–8, July 1999.

[25] M.R. Bai, C.T. Tung, and C.C. Lee, "Subband Approach to Bandlimited Crosstalk Cancellation System in Spatial Sound Reproduction," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, 2007.

[26] M. Kallinger and A. Mertins, "A Spatially Robust Least Squares Crosstalk Canceller," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 177–180, Apr. 2007.

[27] P. Majdak, B. Masiero, and J. Fels, "Sound localization in individualized and non-individualized crosstalk cancellation systems," *Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2055-68, Apr. 2013.

[28] A. Farina, "Simultaneous measurements of impulse response and distortion with a swept-sine technique", *Journal of the Acoustical Society of America*, vol. 48, p. 350, Feb. 2000.