Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers

Edgar Y. Choueiri
Princeton University

choueiri@princeton.edu

Crosstalk cancellation (XTC) yields high-spatial-fidelity reproduction of binaural audio through loudspeakers allowing a listener to perceive an accurate 3-D image of a recorded soundfield. Such accurate 3-D sound reproduction is useful in a wide range of applications in the medical, military and commercial audio sectors. However, XTC is known to add a severe spectral coloration to the sound and that has been an impediment to the wide adoption of loudspeaker-based binaural audio. The nature of this coloration in two-loudspeaker XTC systems, and the fundamental aspects of the regularization methods that can be used to optimally control it, were studied analytically using a free-field two-point-source model. It was shown that constant-parameter regularization, while effective at decreasing coloration peaks, does not yield optimal XTC filters, and can lead to the formation of roll-offs and doublet peaks in the filter's frequency response. Frequency-dependent regularization was shown to be significantly better for XTC optimization, and was used to derive a prescription for designing optimal two-loudspeaker XTC filters, whereby the audio spectrum is divided into adjacent bands, each of is which associated with one of three XTC impulse responses, which were derived analytically. Aside from the sought fundamental insight, the analysis led to the formulation of band-assembled XTC filters, whose optimal properties favor their practical use for enhancing the spatial realism of two-loudspeaker playback of standard stereo recordings containing binaural cues.

I. INTRODUCTION

A. Background and Motivation

The ultimate goal of binaural audio with loudspeakers (BAL), also known as transauralization[1], is to reproduce, at the entrance of each of the listener's ear canals, the sound pressure signals recorded on only the ipsilateral channel of a stereo signal. If the stereo signal [2] was encoded with the head-related transfer function (HRTF) of the listener, and includes the proper ITD (interaural time difference) and ILD (interaural level difference) cues, then delivering the signal on each of the channels of the stereo signal to the ipsilateral ear, and only to that ear, would ideally guarantee that the ear-brain system receives the cues it needs to hear an accurate 3-D reproduction of the recorded soundfield. Since, with playback from two loudspeakers, each of them is also heard by the contralateral ear (crosstalk), approaching the goal of BAL requires an effective cancellation of this unintended crosstalk.

In addition to crosstalk cancellation (XTC), effective BAL requires an abatement of sound reflections in the listening room, which cause degradation to the integrity of the binaural cues at the listener's ears[3–5]. While this problem can be somewhat alleviated through prescriptions that increase the ratio of direct to reflected sound, full disambiguation of front-back sound localization through BAL has been shown to require XTC levels[6] above 20 dB, which are difficult to achieve practically even under anechoic conditions[3].

Therefore, it would seem that the goal stated in the first paragraph could be more naturally reached with binaural audio through headphones (or earphones)[7] as both crosstalk and room reflections would be nonexistent. However, with earphones or headphones, the location of the playback transducers in or very near the ears means that non-idealities, (e.g., mismatches between the HRTF of the listener and that used to encode the recording, movement of the perceived sound image with movement of the listener's head, lack of bone-conducted sound, transducer-induced resonances in the ear canal, discomfort, etc.), when above a certain threshold, can lead to difficulties in perceiving a realistic 3-D image and to the perception that the sound (or some of its spectral components) is inside, or too close to, the listener's head.

Binaural playback through loudspeakers is largely immune to this head internalization of sound, for even when non-idealities in binaural reproduction are present, the sound originates far enough from the listener to be perceived to come from outside the head. Furthermore, cues such as bone-conducted sound and the involvement of the listener's own head, torso and pinnae in sound diffraction and reflection during playback (even if it departs from, or interferes with, the diffraction-induced coloration represented in the HRTF used to encode the binaural recording) could be expected to enhance the perceived realism of sound reproduction relative to that achieved with earphones. These potential advantages have, implicitly or explicitly, motivated the development of XTC-enabled BAL since the earliest work on the subject[8–10].

In scientific applications of BAL, such as its application to study spatial hearing disabilities of elderly adults[3], the high levels of XTC (above 20 dB) needed for highly-accurate transmission of binaural cues to a listener require anechoic, or semi-anechoic, environments, precise matching of the listener's HRTF with that used in the recording, and constrained positioning of the listener's head in the area of equalization ("sweet spot"). In many less stringent applications[10], modest levels of XTC, even of a few dB over a limited range of frequencies, have the potential of significantly enhancing the 3-D realism of the reproduction of recordings containing binaural cues. This is because, by definition, localization cues in a binaural recording represent differential interaural information that is intended to be transmitted to the ears with no crosstalk. In other words, crosstalk cancellation, at any level, is a reduction of unintended artifice in the loudspeaker playback of recordings containing significant binaural cues.

This reduction of unintended artifice through XTC should also apply to the loudspeaker playback of most stereo recordings[11], especially those made in real acoustic spaces, and even to recordings made using standard stereo microphone techniques without a dummy head, since these techniques[12] all rely on preserving in the recording a good measure of the natural ILD and ITD cues needed for spatial localization during playback. We should therefore expect that effecting even a relatively low level of XTC to the playback of such standard stereo recordings, even those lacking HRTF encoding, should enhance image localization compared to playback with full crosstalk, as well as the perception of width and depth of the sound-field, since these binaural features are always, to some degree, corrupted by crosstalk[13].

With such promises of high-spatial-fidelity reproduction of binaural recordings, and enhanced realism to the playback of a large portion of existing acoustic stereo recordings, the question arises as to why crosstalk cancellation has not yet penetrated widely in the professional and consumer audio sectors.

A part of the answer to this question is related to the physical constraints required for the practical implementation of an effective XTC-enabled BAL playback system. These constraints include sensitivity to head movements and a limited sweet spot[14–18], sensitivity to room reflections[4, 5], and the often-required departure from the well-established stereo loudspeaker configuration[19, 20] (where the loudspeakers span an angle of 60 degrees with respect to the listener). Much research effort has been expended recently on relieving some of these constraints and has resulted in potential solutions, of varying degrees of practicality, which include: widening the sweet spot through the use of multiple loudspeakers[21–24], providing XTC at multiple listening locations[25], enhancing robustness to head movement through the use of sum and difference filters[26], and dynamically moving the sweet spot to follow the location of the listener's head by tracking it with optical sensors[27].

Another major impediment to the wide adoption of XTC-enabled BAL has been the *spectral coloration* that XTC filters inherently impose on the sound emitted by the loudspeakers. The fundamental nature of this spectral coloration, its basic features, its dependencies, and optimal methods to abate it with minimal adverse effects on XTC performance, are the main subjects of this paper.

B. The Problem of XTC-induced Spectral Coloration

1. Nature of the Problem

One main difficulty in implementing XTC is to reduce the artifice of crosstalk without adding an artifice of another kind: spectral coloration. Sound waves traveling from two distinct sources to the ears set up an interference pattern in the intervening air space. Depending on the frequency, the distance of the ears from the loudspeakers, the distance between the loudspeakers, and the phase relationship between the left and right components of the recorded stereo signal, the wave interference at an ear of the listener might be destructive, complementary, or constructive. At some of the frequencies for which the interference is destructive at the ear, XTC control (i.e., signal processing that would cause the waves from the loudspeakers to the contralateral ears to be nulled) would require boosting the amplitude of the emitted waves. As we shall see in Section II C, for typical listening configurations these level boosts[28] in the case of a *perfect* XTC filter (defined as one that theoretically yields an infinite XTC level over the entire audio band, in a free-field or anechoic environment) can easily be in excess of 30 dB, and therefore amount to severe spectral coloration.

Of course, such a "perfect" XTC filter would impose these necessary level boosts *only at the loudspeakers* in such a way that, *at the listener's ears*, not only the crosstalk is cancelled, but also the frequency spectrum is reconstructed perfectly, i.e., with no spectral coloration.

As recognized in Ref. [29, 30], and as will be further discussed in Section II C, the frequencies at which the level boosts are required correspond to the frequencies at which system inversion (the mathematical inversion of the system's transfer matrix, which leads to the XTC filter) is ill-conditioned. At these frequencies, XTC control becomes highly sensitive to errors[30], so that even a small error in the alignment of the listener's head in the real world would lead to an effective loss of XTC control at, and near, these frequencies. Therefore, not only would there be undesired crosstalk at the listener's ears at these frequencies, but also, and consequently, the levels boosts which must necessarily be imposed at these frequencies, would be fully hearable, even in the sweet spot, as a coloration.

Even in an ideal world where the loudspeakers-listener alignment is perfect, this spectral coloration imposed at the loudspeakers would present three probems: 1) it would be heard by a listener outside the sweet spot, 2) it would cause a relative increase (compared to unprocessed sound playback) in the physical strain on the playback transducers, and 3) it would correspond to a loss in the dynamic range[29]. Since even professional audio equipment is seldom designed to have more more than a few dB headroom above the levels required to reproduce realistic SPL peaks[31], the dynamic range of the program must be decreased by more than 30 dB (minus the headroom), in the case of the "perfect" XTC filter defined above, to avoid clipping. This is particularly problematic, for instance, in the case of wide-dynamic-range audio recorded in 16 bits.

2. Previous Work and Present Goals

Takeuchi and Nelson[29] have developed a method that not only yields excellent measured XTC performance[3, 32], but also effectively solves the problem of spectral coloration. However, their method, called "Optimal Source Distribution" (OSD), which is discussed in Section II C.4, requires the use of a minimum of six transducers positioned at various angles around the listener.

The problem of XTC-induced spectral coloration for playback with only two loudspeakers remains compelling due to the simplicity of the two-loudspeaker set-up and its compatibility with existing audio equipment. In this paper, we study this problem in the context of XTC optimization, which we define as maximizing XTC performance for a desired tolerable level of spectral coloration or, equivalently, minimizing the spectral coloration for a desired XTC performance.

In particular, we use a free-field two-point-source model and address, analytically, the fundamental aspects of spectral coloration control through both constant-parameter and frequency-dependent regularization methods. While regularization methods have been used in the audio literature for sound source reconstruction[33] and to control ill-conditioning in HRTF inversion[3, 23, 24, 34] their fundamental properties in the context of XTC optimization and XTC-induced spectral coloration have not been studied in detail, especially in the case of frequency-dependent regularization.

Aside from the fundamental insight we seek through this analysis, we aim to derive analytical expressions for the time-domain impulse responses (IRs) of such optimal XTC filters. These IRs are not only useful for shedding light on the time-domain aspects of XTC optimization, but can also lend their benefits, through digital convolution engines, to (typically non-scientific) applications where the needed XTC levels are below those requiring inversion of the listener's HRTF, and are sufficient to enhance the spatial realism of the two-loudspeaker playback (in regular listening rooms) of audio signals containing significant binaural cues.

II. THE FUNDAMENTAL XTC PROBLEM

In this section we start with the mathematical formulation of the model and the governing transformation matrices. We then define a set of metrics that will be useful for evaluating and comparing the spectral coloration and performance of XTC filters, and conclude with the definition and discussion of a benchmark for such comparisons: the perfect XTC filter.

A. Formulation and Transformation Matrices

In order to render the analysis tractable enough so that fundamental insight is more easily obtained, we make the idealizing assumptions that sound propagation occurs in a free field (with no diffraction or refelection from the head and pinnae of the listener or any other physical objects), and that the loudspeakers radiate like point sources.

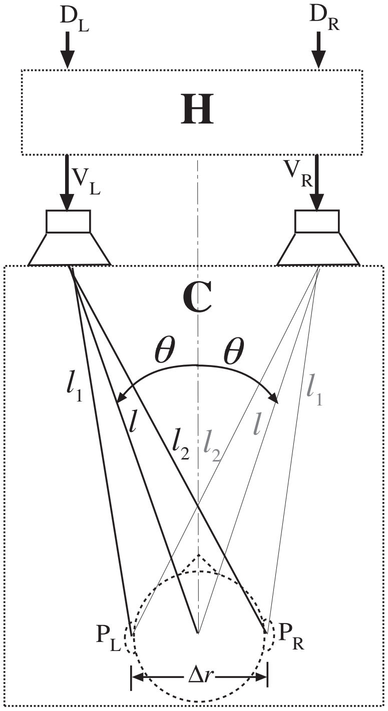

FIG. 1: Geometry of the two-source free-field model. (All symbols are defined in the text.)

In the frequency domain, the air pressure at a free-field point located a distance r from a point source (monopole) radiating a sound wave of frequency ω is given by[35]

$$P(r, i\omega) = \frac{i\omega\rho_o q}{4\pi}\frac{e^{-ikr}}{r},$$

where ρ_0 is the air density, $k = 2\pi/\lambda = \omega/c_s$ the wavenumber, λ the wavelength, c_s the speed of sound (340.3 m/s), and q the source strength (in units of volume per unit time). It is convenient to define

$$V = \frac{i\omega\rho_0 q}{4\pi},$$

which is the time derivative of $\rho_o q/(4\pi)$, the mass flow rate of air from the center of the source.

Therefore, at the left ear of a listener in the symmetric two-source geometry shown in Fig. 1, the air pressure due to the two sources, under the above-stated assumptions, add up as

$$P_L(i\omega) = \frac{e^{-ikl_1}}{l_1}V_L(i\omega) + \frac{e^{-ikl_2}}{l_2}V_R(i\omega). \quad (1)$$

Similarly, at the right ear, we have

$$P_R(i\omega) = \frac{e^{-ikl_2}}{l_2}V_L(i\omega) + \frac{e^{-ikl_1}}{l_1}V_R(i\omega). \quad (2)$$

Here, l_1 and l_2 are the path lengths between any of the two sources and the ipsilateral and contralateral ear, respectively, as shown in that figure.

In order to maintain a connection with the relevant literature, we adopt the same nomenclature used in Refs. [19, 20, 29, 30]. Namely, unless otherwise stated, we use uppercase letters for frequency variables, lowercase for time-domain variables, uppercase bold for matrices and lowercase bold for vectors, and define

$$\Delta l \equiv l_2 - l_1 \quad \text{and} \quad g \equiv l_1/l_2 \quad (3)$$

as the path length difference and path length ratio, respectively. An inspection of the geometry illustrated in Fig. 1 shows that $0 < g < 1$, and that the path lengths can be expressed as

$$l_1 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 - \Delta r\, l \sin(\theta)}, \quad (4)$$

$$l_2 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 + \Delta r\, l \sin(\theta)}, \quad (5)$$

where Δr is the effective distance between the entrances of the ear canals, and l is the distance between either source and the interaural mid-point. As defined in Fig. 1, $\Theta = 2\theta$ is the loudspeaker span. Note that for $l >> \Delta r \sin(\theta)$, as in most loudspeaker-based listening set-ups, we have $g \simeq 1$. Another important parameter is the time delay,

$$\tau_c = \frac{\Delta l}{c_s}, \quad (6)$$

defined as the time it takes a sound wave to traverse the path length difference Δl.

Using the above definitions, Eqs. (1) and (2) can be re-written in matrix form as

$$\begin{bmatrix} P_L(i\omega) \\ P_R(i\omega) \end{bmatrix} = \alpha \begin{bmatrix} 1 & ge^{-i\omega\tau_c} \\ ge^{-i\omega\tau_c} & 1 \end{bmatrix} \begin{bmatrix} V_L(i\omega) \\ V_R(i\omega) \end{bmatrix}, \quad (7)$$

where

$$\alpha = \frac{e^{-i\omega l_1/c_s}}{l_1}. \quad (8)$$

In the time domain, α is simply a transmission delay (divided by the constant l_1) that does not affect the shape of the signal. Its role in insuring causality is discussed in Section III B. The source vector $\boldsymbol{v} = [V_L(i\omega), V_R(i\omega)]^T$ is obtained from the vector of "recorded" signals $\boldsymbol{d} = [D_L(i\omega), D_R(i\omega)]^T$, through the transformation

$$\boldsymbol{v} = \boldsymbol{Hd}, \quad (9)$$

where

$$\boldsymbol{H} = \begin{bmatrix} H_{LL}(i\omega) & H_{LR}(i\omega) \\ H_{RL}(i\omega) & H_{RR}(i\omega) \end{bmatrix} \quad (10)$$

is the sought 2×2 filter matrix. Therefore, from Eq. (7), we have

$$\boldsymbol{p} = \alpha \boldsymbol{CHd}, \quad (11)$$

where $\boldsymbol{p} = [P_L(i\omega), P_R(i\omega)]^T$ is the vector of pressures at the ears, and $\boldsymbol{C}$ is the system's transfer matrix

$$\boldsymbol{C} \equiv \begin{bmatrix} 1 & ge^{-i\omega\tau_c} \\ ge^{-i\omega\tau_c} & 1 \end{bmatrix}, \quad (12)$$

which, like all matrices we will be dealing with, is symmetric due to the symmetry of the geometry.

In summary, the transformation from the signal $\boldsymbol{d}$, through the filter $\boldsymbol{H}$, to the source variables $\boldsymbol{v}$, then through wave propagation from the sources to pressure $\boldsymbol{p}$ at the ears of the listener, can be written simply as

$$\boldsymbol{p} = \alpha \boldsymbol{Rd}. \quad (13)$$

where we have introduced the performance matrix, $\boldsymbol{R}$, defined as

$$\boldsymbol{R} = \begin{bmatrix} R_{LL}(i\omega) & R_{LR}(i\omega) \\ R_{RL}(i\omega) & R_{RR}(i\omega) \end{bmatrix} \equiv CH. \quad (14)$$

B. Metrics

We now wish to define a set of metrics by which to judge the spectral coloration and performance of XTC filters. In this context we note that the diagonal elements of $\boldsymbol{R}$ represent the ipsilateral transmission of the signal to the ears, and the off-diagonal elements represent the undesired contralateral transmission, i.e., the crosstalk.

Therefore, the amplitude spectrum (to a factor α) of a signal fed to only one (either left or right) of the two inputs of the system, as heard at the ipsilateral ear is

$$E_{\text{si}_{\|}}(\omega) \equiv |R_{LL}(i\omega)| = |R_{RR}(i\omega)|,$$

where the subscripts "si" and $\|$ stand for "side image" and "ipsilateral ear (with respect to the input signal)" respectively, since $E_{\text{si}_{ip}}$, as defined, is the frequency response (at the ipsilateral ear) for the side image that would result from the input being panned to one side. Similarly, at the contralateral ear to the input signal (subscript X), we have the following side-image frequency response:

$$E_{\mathrm{si}_X}(\omega) \equiv |R_{LR}(i\omega)| = |R_{RL}(i\omega)|.$$

The system's frequency response at either ear when the same signal is split equally between left and right inputs is another spectral coloration metric. It can be obtained from the product $\boldsymbol{R} \cdot [1/2, 1/2]^T$, which leads to

$$E_{\mathrm{ci}}(\omega) \equiv |\frac{R_{LL}(i\omega) + R_{LR}(i\omega)}{2}| = |\frac{R_{RL}(i\omega) + R_{RR}(i\omega)}{2}|.$$

Here the subscript “ci” stands for “center image” since E_{ci}, as defined, is the frequency response (at either ear) for the center image that would result from the input being panned to the center.

Also of importance to our discussions are the frequency responses that would be measured at the sources (loudspeakers). These are denoted by S, and can be obtained from the elements of the filter matrix $\boldsymbol{H}$. They are given using the same subscript convention used above (with “$\|$” and “X” referring to the loudspeakers that are ipsilateral and contralateral to the input signal, respectively) by

$$S_{\mathrm{si}_\|}(\omega) \equiv |H_{LL}(i\omega)| = |H_{RR}(i\omega)|,$$

$$S_{\mathrm{si}_X}(\omega) \equiv |H_{LR}(i\omega)| = |H_{RL}(i\omega)|,$$

$$S_{\mathrm{ci}}(\omega) \equiv |\frac{H_{LL}(i\omega) + H_{LR}(i\omega)}{2}| = |\frac{H_{RL}(i\omega) + H_{RR}(i\omega)}{2}|.$$

An intuitive interpretation of the significance of the above metrics is that a signal panned from a single input to both inputs to the system will result in frequency responses going from E_{si} to E_{ci} at the ears, and S_{si} to S_{ci} at the loudspeakers.

Two other spectral coloration metrics are the frequency responses of the system to in-phase and out-of-phase inputs to the system. These two responses are obtained simply from the product of the filter matrix $\boldsymbol{H}$ with the vectors $[1,1]^T$ and $[1,-1]^T$ (or $[-1,1]^T$), respectively, and are given by:

$$\begin{aligned} S_i(\omega) &\equiv |H_{LL}(i\omega) + H_{LR}(i\omega)| = |H_{RL}(i\omega) + H_{RR}(i\omega)|, \\ S_o(\omega) &\equiv |H_{LL}(i\omega) - H_{LR}(i\omega)| = |H_{RL}(i\omega) - H_{RR}(i\omega)|, \end{aligned}$$

where the subscripts i and o denote the in-phase and out-of-phase responses, respectively. Note that, as defined, S_i is double (i.e., 6 dB above) S_{ci}, as the latter describes a signal of amplitude 1 panned to center (i.e., split equally between L and R inputs), while the former describes two signals of amplitude 1 fed in phase to the two inputs of the system.

Since a real signal can consist of various components having different phase relationships, it is more useful to combine $S_i(\omega)$ and $S_o(\omega)$ into a single metric, $\hat{S}(\omega)$, which is the *envelope spectrum* that describes the maximum amplitude that could be expected at the loudspeakers, and is given by

$$\hat{S}(\omega) = \max\left[S_i(\omega), S_o(\omega)\right].$$

It is relevant to note that $\hat{S}(\omega)$ is equivalent to $||\boldsymbol{H}||$, the 2-norm of $\boldsymbol{H}$, and that S_i and S_o are the two singular values, which can be obtained through singular value decomposition of the matrix as was done in Ref. [29].

Finally, an important metric that will allow us to evaluate and compare the XTC performance of various filters is $\chi(\omega)$, the crosstalk cancellation spectrum:

$$\chi(\omega) \equiv \frac{|R_{LL}(i\omega)|}{|R_{RL}(i\omega)|} = \frac{|R_{RR}(i\omega)|}{|R_{LR}(i\omega)|} = \frac{E_{\mathrm{si}_\|}(\omega)}{E_{\mathrm{si}_X}(\omega)}.$$

The above definitions give us a total of eight metrics, ($E_{\mathrm{si}_\|}$, E_{si_X}, E_{ci}, $S_{\mathrm{si}_\|}$, S_{si_X}, S_{ci}, $\hat{S}, \chi$), all real functions of frequency, by which to evaluate and compare the spectral coloration and XTC performance of XTC filters.

C. Benchmark: Perfect Crosstalk Cancellation

A perfect crosstalk cancellation (P-XTC) filter is defined as one that, theoretically, yields infinite crosstalk cancellation at the ears of the listener, for all frequencies.

Crosstalk cancellation, as defined in Section I A, requires that the pressure at each of the two ears be that which would have resulted from the ipsilateral signal alone, namely, in the frequency domain, $P_L = \alpha D_L$ and $P_R = \alpha D_R$, where all quantities are complex functions of frequency. Therefore, in order to achieve perfect cancellation of the crosstalk, Eq. (13) requires that $\boldsymbol{R} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the unity matrix, and thus, as per the definition of $\boldsymbol{R}$ in Eq. (14), the P-XTC filter is simply the inverse of the system transfer matrix expressed in Eq. (12), and can be expressed exactly:

$$\boldsymbol{H}^{[P]} = \boldsymbol{C}^{-1} = \frac{1}{1 - g^2 e^{-2i\omega\tau_c}} \begin{bmatrix} 1 & -ge^{-i\omega\tau_c} \\ -ge^{-i\omega\tau_c} & 1 \end{bmatrix}, \tag{15}$$

where the superscript $[P]$ denotes perfect XTC. For this filter, the eight metrics we defined above become:

$$\begin{aligned} E^{[P]}_{\mathrm{si}_\|} &= 1; \qquad E^{[P]}_{\mathrm{si}_X} = 0; \qquad E^{[P]}_{\mathrm{ci}} = \frac{1}{2}; \\ S^{[P]}_{\mathrm{si}_\|}(\omega) &= \left|\frac{1}{1 - g^2 e^{-2i\omega\tau_c}}\right| = \frac{1}{\sqrt{g^4 - 2g^2\cos(2\omega\tau_c) + 1}}; \\ S^{[P]}_{\mathrm{si}_X}(\omega) &= \left|\frac{-ge^{-i\omega\tau_c}}{1 - g^2 e^{-2i\omega\tau_c}}\right| = \frac{g}{\sqrt{g^4 - 2g^2\cos(2\omega\tau_c) + 1}}; \\ S^{[P]}_{\mathrm{ci}}(\omega) &= \frac{1}{2}\left|1 - \frac{g}{g + e^{i\omega\tau_c}}\right| = \frac{1}{2\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}; \\ \hat{S}^{[P]}(\omega) &= \max\left(\left|1 - \frac{g}{g + e^{i\omega\tau_c}}\right|, \left|1 + \frac{g}{e^{i\omega\tau_c} - g}\right|\right), \end{aligned}$$

$$= \max\left(\frac{1}{\sqrt{g^2+2g\cos(\omega\tau_c)+1}},\right.$$

$$\left.\frac{1}{\sqrt{g^2-2g\cos(\omega\tau_c)+1}}\right); \quad (16)$$

$$\chi^{[P]}(\omega) = \infty. \quad (17)$$

Therefore the perfect ($\chi = \infty$) XTC filter gives flat frequency responses at the ears ($E^{[P]}(\omega) =$ constant), but not at the sources. To appreciate the extent of spectral coloration at the loudspeakers, we plot the $S^{[P]}(\omega)$ frequency responses expressed above in Fig. 2 for a typical value $g = .985$. Throughout this paper, for the sake of illustration, we complement the non-dimensional plots with dimensional calculations, which are represented by the same curves read in terms of the frequency $f = \omega/2\pi$ on the top axis, for a typical listening geometry characterized by $g = .985$ and $\tau_c = 68$ μs (i.e., 3 samples at the Red Book CD sampling rate of 44.1kHz), which would be the case, for instance, of a set-up with $\Delta r = 15$ cm, $l = 1.6$ m, and $\Theta = 18°$.)

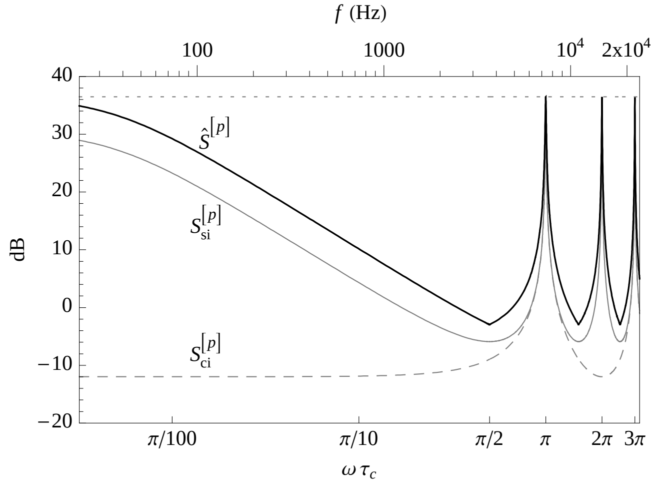

FIG. 2: Perfect XTC filter frequency responses at the loudspeakers: amplitude envelope (heavy curve), side image (light solid curve), and central image (light dashed curve). The dotted horizontal line marks the envelope ceiling, which for this case ($g = .985$) is 36.5 dB. The non-dimensional frequency $\omega\tau_c$ is given on the bottom axis, and the corresponding frequency in Hz, shown on the top axis, is to illustrate a particular (typical) case of τ_c =3 samples at a sampling rate of 44.1 kHz. (Since $S^{[P]}_{\text{si}_\parallel} \simeq S^{[P]}_{\text{si}_X}$ when $g \simeq 1$, these two spectra are shown as the single curve $S^{[P]}_{\text{si}}$.)

The peaks in these spectra occur at frequencies for which the system must boost the amplitude of the signal at the loudspeakers in order to effect XTC at the ears while compensating for the destructive interference at that location. Similarly, minima in the spectra occur when the amplitude must be attenuated.

Using the first and second derivatives (with respect to $\omega\tau_c$) of the above expressions for the various $S^{[P]}(\omega)$ spectra, we find the following amplitudes and frequencies for the associated peaks and minima, denoted by ↑ and ↓ superscripts, respectively:

$$\begin{aligned}
S^{[P]^\uparrow}_{\text{si}_\parallel} &= \frac{1}{1-g^2} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots\\
S^{[P]^\downarrow}_{\text{si}_\parallel} &= \frac{1}{1+g^2} \text{ at } \omega\tau_c = n\frac{\pi}{2}, \text{ with } n = 1,3,5,7,\ldots\\
S^{[P]^\uparrow}_{\text{si}_X} &= \frac{g}{1-g^2} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots\\
S^{[P]^\downarrow}_{\text{si}_X} &= \frac{g}{1+g^2} \text{ at } \omega\tau_c = n\frac{\pi}{2}, \text{ with } n = 1,3,5,7,\ldots\\
S^{[P]^\uparrow}_{\text{ci}} &= \frac{1}{2-2g} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 1,3,5,7,\ldots\\
S^{[P]^\downarrow}_{\text{ci}} &= \frac{1}{2+2g} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 0,2,4,6,\ldots\\
\hat{S}^{[P]^\uparrow} &= \frac{1}{1-g} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots
\end{aligned} \quad (18)$$

$$\hat{S}^{[P]^\downarrow} = \frac{1}{\sqrt{1+g^2}} \text{ at } \omega\tau_c = n\frac{\pi}{2}, \text{ with } n = 1,3,5,7,\ldots \quad (19)$$

For a typical listening set-up, $g \simeq 1$, say, our reference $g = .985$ case shown in Fig. 2, the envelope peaks (i.e., $\hat{S}^{[P]^\uparrow}$) correspond to a boost of

$$20\log_{10}\left(\frac{1}{1-.985}\right) = 36.5 \text{ dB}$$

(and the peaks in the other spectra, $S^{[P]^\uparrow}_{\text{si}_\parallel} \simeq S^{[P]^\uparrow}_{\text{si}_X} \simeq S^{[P]^\uparrow}_{\text{ci}}$, correspond to boosts of about 30.5 dB.) While these boosts have equal frequency widths across the spectrum, when the spectrum is plotted logarithmically (as is appropriate for human sound perception), the low-frequency boost is most prominent in its perceived frequency extent. This bass boost has long been recognized as an intrinsic problem in XTC. While the high-frequency peaks could, in principle, be pushed out of the audio range by decreasing τ_c (which, as can be seen from Eqs. (4) to (6), is achieved by increasing l and/or decreasing the loudspeaker span Θ, as is done in the so-called "Stereo Dipole" configuration described in Ref. [19, 20], where $\Theta = 10°$), the "low frequency boost" of the P-XTC filter would remain problematic.

As mentioned in Section IB1, the severe spectral coloration associated with these high-amplitude peaks presents three practical problems: 1) it would be heard by a listener outside the sweet spot, 2) it would cause a relative increase (compared to unprocessed sound playback) in the physical strain on the playback transducers, and 3) it would correspond to a loss in the dynamic range.

These penalties might be a justifiable price to pay if we are guaranteed the infinitely good XTC performance ($\chi = \infty$) and the perfectly flat frequency response ($E^{[P]}(\omega) = \text{constant}$) that the perfect XTC filter promises at the ears of a listener in the sweet spot. However, in practice, these theoretically promised benefits are unachievable due to the solution's sensitivity to unavoidable errors. This problem can best be appreciated by evaluating the condition number of the transfer matrix $\boldsymbol{C}$.

It is well known that in matrix inversion problems the sensitivity of the solution to errors in the system is given by the condition number of the matrix. (For a discussion of the condition number in the context of XTC system errors, see Ref. [30]). The condition number $\kappa(\boldsymbol{C})$ of the matrix $\boldsymbol{C}$ is given by

$$\kappa(\boldsymbol{C}) = ||\boldsymbol{C}||\;||\boldsymbol{C}^{-1}|| = ||\boldsymbol{C}||\;||\boldsymbol{H}^{[P]}||.$$

(It is also, equivalently, the ratio of largest to smallest singular values of the matrix.) Therefore, we have

$$\begin{aligned}\kappa(\boldsymbol{C}) \;=\; \max\Bigg(&\sqrt{\frac{2(g^2+1)}{g^2+2g\cos(\omega\tau_c)+1}-1},\\ &\sqrt{\frac{2(g^2+1)}{g^2-2g\cos(\omega\tau_c)+1}-1}\Bigg).\end{aligned}$$

Using the first and second derivatives of this function, as we did for the previous spectra, we find the following maxima and minima:

$$\begin{aligned}\kappa^{\uparrow}(\boldsymbol{C}) \;&=\; \frac{1+g}{1-g} \text{ at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots\\ \kappa^{\downarrow}(\boldsymbol{C}) \;&=\; 1 \quad \text{at } \omega\tau_c = n\frac{\pi}{2}, \text{ with } n = 1,3,5,7,\ldots \qquad (20)\end{aligned}$$

as was also reported in Ref. [30] in terms of wavelengths. First, we note that the peaks and minima in the condition number occur at the same frequencies as those of the amplitude envelope spectrum at the loudspeakers, $\hat{S}^{[P]}$. Second, we note that the minima have a condition number of unity (the lowest possible value), which implies that the filter resulting from the inversion of $\boldsymbol{C}$ is most robust (i.e., least sensitive to errors in the transfer matrix) at the non-dimensional frequencies $\omega\tau_c = \pi/2, 3\pi/2, 5\pi/2, \ldots$. Conversely, the condition number can reach very high values (e.g., $\kappa^{\uparrow}(\boldsymbol{C}) = 132.3$ for our typical case of $g = .985$) at the non-dimensional frequencies $\omega\tau_c = 0, \pi, 2\pi, 3\pi \ldots$. As $g \to 1$ the matrix inversion resulting in the P-XTC filter becomes ill-conditioned, or in other words, infinitely sensitive to errors. The slightest misalignment, for instance, of the listener's head, would thus result in a severe loss in XTC control at the ears (at and near these frequencies) which, in turn, causes the severe spectral coloration in $\hat{S}^{[P]}(\omega)$ to be transmitted to the ears.

We are now in a position to appreciate the prescription proposed and implemented by Takeuchi and Nelson[29, 32], which effectively solves both the robustness and spectral coloration problem of the P-XTC filter by insuring that the system operates always under conditions where $\kappa(\boldsymbol{C})$ is small. This can be done by allowing the loudspeaker span to be a function of the frequency. More specifically, after noting that typically $l >> \Delta r$, so that the approximation $\Delta l \simeq \Delta r \sin(\theta)$ holds, and therefore $\omega\tau_c = \omega\Delta l/c_s = 2\pi f\Delta l/c_s$ can be approximated by

$$\omega\tau_c \simeq \frac{2\pi f\Delta r\sin(\theta)}{c_s} \qquad \text{for } l \gg \Delta r, \qquad (21)$$

we can re-write the robustness condition (stated in Eq. (20)) as

$$\Theta(f) \simeq 2\sin^{-1}\left(\frac{nc_s}{4f\Delta r}\right), \quad \text{with } n = 1,3,5,7,\ldots$$

Since both c_s and Δr are constant, the required loudspeaker span is solely a function of the frequency f. In practice this prescription, called Optimal Source Distribution (OSD), can be implemented by using a crossover network to distribute adjacent bands of the audio spectrum to pairs of transducers, whose spans are calculated from the above equation so that in each band the condition number does not exceed unity by much, thus insuring robustness and low coloration over the entire audio spectrum. It is clear, however, that this solution is not applicable to the case of a single pair of loudspeakers, which is the focus of our analysis.

We refer the reader interested in the OSD method and XTC errors to Ref. [29, 30, 32], and sum up the discussion in this section by stating that, for the case of only two loudspeakers, the perfect XTC filter carries in practice the penalties of over-amplification (and the associated loss of dynamic range) at frequencies where system inversion is ill-conditioned, transducer fatigue, and a severe spectral coloration that is heard by listeners inside and outside the sweet spot.

III. CONSTANT-PARAMETER REGULARIZATION

Regularization methods allow controlling the norm of the approximate solution of an ill-conditioned linear system at the price of some loss in the accuracy of the solution. The control of the norm through regularization can be done subject to an optimization prescription, such as the minimization of a cost function. Ref. [36] provides a detailed discussion of regularization methods in a general mathematical context, and Refs. [3, 18, 23, 33, 34] are examples of the use of regularization to control numerical HRTF inversion. We discuss regularization analytically in the context of XTC filter optimization, which we define as the maximization of XTC performance for a desired tolerable level of spectral coloration or, equivalently, the minimization of spectral coloration for a desired minimum XTC performance.

In essence, a nearby solution to the matrix inversion problem is sought:

$$\boldsymbol{H}^{[\beta]} = \left[\boldsymbol{C}^H\boldsymbol{C} + \beta\boldsymbol{I}\right]^{-1}\boldsymbol{C}^H, \qquad (22)$$

where the superscript H denotes the Hermitian operator, and β is the regularization parameter which essentially causes a departure from $\boldsymbol{H}^{[P]}$, the exact inverse of $\boldsymbol{C}$. In this section we take β to be a constant, $0 < \beta \ll 1$. The pseudoinverse matrix $\boldsymbol{H}^{[\beta]}$ is the regularized filter, and the superscript $[\beta]$ is used to denote constant-parameter regularization. The regularization stated in Eq. (22) can be shown[23, 34, 37] to correspond to a minimization of a cost function, $J(i\omega)$,

$$J(i\omega) = \boldsymbol{e}^H(i\omega)\boldsymbol{e}(i\omega) + \beta \boldsymbol{v}^H(i\omega)\boldsymbol{v}(i\omega), \qquad (23)$$

where the vector $\boldsymbol{e}$ represents a performance metric that is a measure of the departure from the signal reproduced by the perfect filter. Physically, then, the first term in the sum constituting the cost function represents a measure of the performance error, and the second term represents an "effort penalty," which is a measure of the power exerted by the loudspeakers. For $\beta > 0$, Eq. (22) leads to an optimum, which corresponds to the least-square minimization of the cost function $J(i\omega)$.

Therefore, an increase of the regularization parameter β leads to a minimization of the effort penalty at the expense of a larger performance error and thus to an abatement of the peaks in the norm of $\boldsymbol{H}$, i.e., the coloration peaks in the $\boldsymbol{S}(\omega)$ spectra, at the price of a decrease in XTC performance at and near the frequencies where the system is ill-conditioned.

A. Frequency Response

Using the explicit form for $\boldsymbol{C}$ given by Eq. (12), in the last equation above, we find:

$$\boldsymbol{H}^{[\beta]} = \begin{bmatrix} H_{LL}^{[\beta]}(i\omega) & H_{LR}^{[\beta]}(i\omega) \\ H_{RL}^{[\beta]}(i\omega) & H_{RR}^{[\beta]}(i\omega) \end{bmatrix}, \qquad (24)$$

where

$$\begin{aligned} H_{LL}^{[\beta]}(i\omega) &= H_{RR}^{[\beta]}(i\omega) \\ &= \frac{g^2 e^{i4\omega\tau_c} - (\beta+1)e^{i2\omega\tau_c}}{g^2 e^{i4\omega\tau_c} + g^2 - [(g^2+\beta)^2 + 2\beta + 1]}, \end{aligned} \qquad (25)$$

$$\begin{aligned} H_{LR}^{[\beta]}(i\omega) &= H_{RL}^{[\beta]}(i\omega) \\ &= \frac{g e^{i\omega\tau_c} - g(g^2+\beta)e^{i3\omega\tau_c}}{g^2 e^{i4\omega\tau_c} + g^2 - [(g^2+\beta)^2 + 2\beta + 1]}. \end{aligned} \qquad (26)$$

The eight metric spectra we defined in Section II B become:

$$\begin{aligned}
E_{\mathrm{si}_\parallel}^{[\beta]}(\omega) &= \frac{g^4 + \beta g^2 - 2g^2\cos(2\omega\tau_c) + \beta + 1}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\
E_{\mathrm{si}_X}^{[\beta]}(\omega) &= \frac{2g\beta|\cos(\omega\tau_c)|}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\
E_{\mathrm{ci}}^{[\beta]}(\omega) &= \frac{1}{2} - \frac{\beta}{2\left[g^2 + 2\cos(\omega\tau_c) + \beta + 1\right]}; \\
S_{\mathrm{si}_\parallel}^{[\beta]}(\omega) &= \frac{\sqrt{g^4 - 2(\beta+1)g^2\cos(2\omega\tau_c) + (\beta+1)^2}}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\
S_{\mathrm{si}_X}^{[\beta]}(\omega) &= \frac{g\sqrt{(g^2+\beta)^2 - 2(g^2+\beta)\cos(2\omega\tau_c) + 1}}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\
S_{\mathrm{ci}}^{[\beta]}(\omega) &= \frac{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}{2[g^2 + 2g\cos(\omega\tau_c) + \beta + 1]}; \\
\hat{S}^{[\beta]}(\omega) &= \max\left(\frac{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}{g^2 + 2g\cos(\omega\tau_c) + \beta + 1}, \right. \\
&\qquad \left. \frac{\sqrt{g^2 - 2g\cos(\omega\tau_c) + 1}}{g^2 - 2g\cos(\omega\tau_c) + \beta + 1}\right); \qquad (27) \\
\chi^{[\beta]}(\omega) &= \frac{g^4 + \beta g^2 - 2g^2\cos(2\omega\tau_c) + \beta + 1}{2g\beta|\cos(\omega\tau_c)|}. \qquad (28)
\end{aligned}$$

Of course, as $\beta \to 0$, $\boldsymbol{H}^{[\beta]} \to \boldsymbol{H}^{[P]}$, and it can be verified that the spectra of the perfect XTC filter are recovered from the expressions above.

The envelope spectrum, $\hat{S}^{[\beta]}(\omega)$, is plotted in Fig. 3 for three values of β. Two features can be noted in that plot: 1) increasing the regularization parameter attenuates the peaks in the spectrum without affecting the minima, and 2) with increasing β the spectral maxima split into doublet peaks (two closely-spaced peaks).

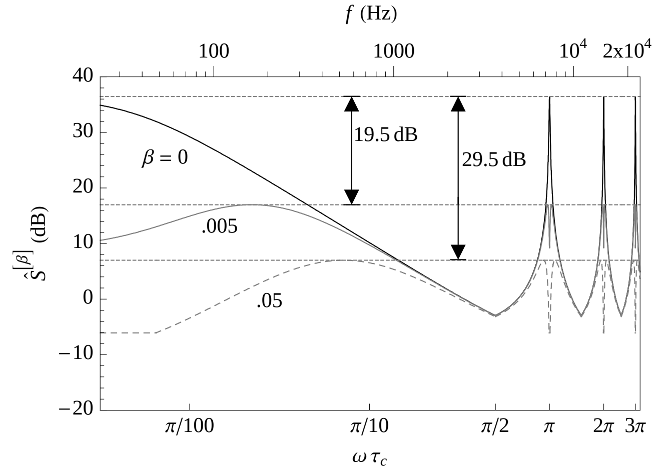

FIG. 3: Effects of regularization on the envelope spectrum at the loudspeakers, $\hat{S}^{[\beta]}(\omega)$, showing peak attenuation and formation of doublet peaks as β is increased. (Other parameters are the same as for Fig. 2.)

To get a measure of peak attenuation and the conditions for the formation of doublet peaks, we take the first and second derivatives of $\hat{S}^{[\beta]}(\omega)$ with respect to $\omega\tau_c$ and find the conditions for which the first derivative is nil and the second is negative. These conditions are summarized as follows: If β is below a threshold β^* defined as

$$\beta < \beta^* \equiv (g-1)^2, \qquad (29)$$

the peaks are singlets and occur at the same nondimensional frequencies as for the envelope spectrum peaks of the P-XTC filter ($\hat{S}^{[P]^{\uparrow}}$), and have the following amplitude:

$$\hat{S}^{[\beta]^{\uparrow}} = \frac{1-g}{(g-1)^2+\beta} \quad \text{at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots$$

If the condition

$$\beta^* \leq \beta \ll 1 \tag{30}$$

is satisfied, the maxima are doublet peaks located at the following non-dimensional frequencies:

$$\omega\tau_c = n\pi \pm \cos^{-1}\left(\frac{g^2-\beta+1}{2g}\right) \text{ with } n = 0,1,2,3,4,\ldots \tag{31}$$

and have an amplitude

$$\hat{S}^{[\beta]^{\uparrow\uparrow}} = \frac{1}{2\sqrt{\beta}}, \tag{32}$$

which does not depend on g. (The superscripts ↑ and ↑↑ denote singlet and doublet peaks, respectively.) The attenuation of peaks in the $\hat{S}^{[\beta]}$ spectrum due to regularization can be obtained by dividing the amplitude of the peaks in the P-XTC (i.e., $\beta = 0$) spectrum by that of peaks in the regularized spectrum. For the case of singlet peaks, the attenuation is

$$20\log_{10}\left(\frac{\hat{S}^{[P]^{\uparrow}}}{\hat{S}^{[\beta]^{\uparrow}}}\right) = 20\log_{10}\left[\frac{\beta}{(g-1)^2+1}\right] \text{ dB},$$

and for doublet peaks, it is given by

$$20\log_{10}\left(\frac{\hat{S}^{[P]^{\uparrow}}}{\hat{S}^{[\beta]^{\uparrow\uparrow}}}\right) = 20\log_{10}\left[\frac{2\sqrt{\beta}}{1-g}\right] \text{ dB}.$$

For the typical case of $g = .985$ illustrated in Fig. 3, we have $\beta^* = 2.225 \times 10^{-4}$, and for $\beta = .005$ and 0.05 we get doublet peaks that are attenuated (with respect to the peaks in the P-XTC spectrum) by 19.5 and 29.5 dB, respectively, as marked on that plot.

Therefore, increasing the regularization parameter above this (typically low) threshold causes the maxima in the envelope spectrum to split into doublet peaks shifted by a frequency $\Delta(\omega\tau_c) = \cos^{-1}[(g^2-\beta+1)/2g]$ to either side of the peaks in the response of the perfect XTC filter. (For our illustrative case of $g = .985$, we have $\beta^* = 2.225 \times 10^{-4}$ and $\Delta(\omega\tau_c) \simeq 0.225$ for $\beta = .05$). Due to the logarithmic nature of frequency perception for humans, these doublet peaks are perceived as narrow-band artifacts at high frequencies (i.e., for $n = 1,2,3,\ldots$), but the first doublet peak centered at $n = 0$ is perceived as a wide-band low-frequency rolloff of typically many dB, as can be clearly seen in Fig. 3. Therefore, constant-β regularization transforms the bass boost of the perfect XTC filter into a bass roll-off.

Since regularization is essentially a deliberate introduction of error into system inversion, we should expect both the XTC spectrum and the frequency responses at the ears to suffer (i.e., depart from their ideal P-XTC filter levels of ∞ and 0 dB, respectively) with increasing β. The effects of constant-parameter regularization on responses at the ears are illustrated in Fig. 4.

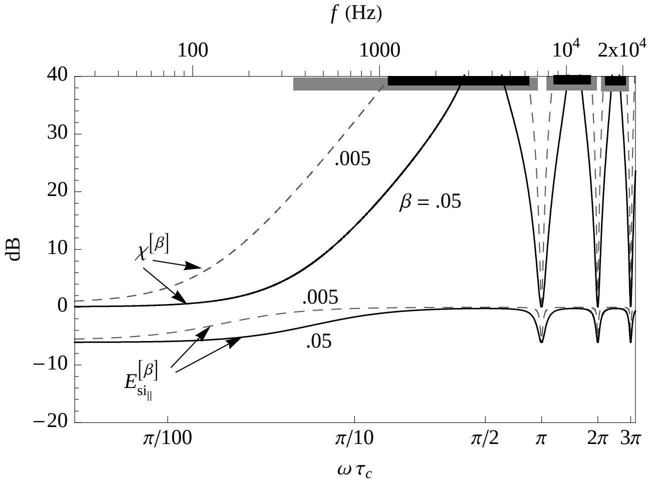

FIG. 4: Effects of regularization on the the crosstalk cancellation spectrum, $\chi^{[\beta]}(\omega)$ (top two curves), and the ipsilateral frequency response at the ear for a side image, $E^{[\beta]}_{\text{si}_\|}(\omega)$ (bottom two curves). The black horizontal bars on the top axis mark the frequency ranges for which an XTC level of 20 dB or higher is reached with $\beta = .05$, and the grey bars represent the same for the case of $\beta = .005$. (Other parameters are the same as for Fig. 2.)

The black curves in that plot represent the crosstalk cancellation spectra and show that XTC control is lost within frequency bands centered around the frequencies where the system is ill-conditioned ($\omega\tau_c = n\pi$ with $n = 0,1,2,3,4,\ldots$) and whose frequency extent widens with increasing regularization. For example, increasing β to .05 limits XTC of 20 dB or higher to the frequency ranges marked by black horizontal bars on the top axis of that figure, with the first range extending only from 1.1 to 6.3 kHz and the second and third ranges located above 8.4 kHz. In many practical applications, such high (20 dB) XTC levels may not be needed or achievable (e.g., because of room reflections and/or HRTF mismatch) and the higher values of β needed to tame the spectral coloration peaks below a required level at the loudspeakers may be tolerated.

The $E^{[\beta]}_{\text{si}_\|}(\omega)$ responses at the ears, shown as the bottom curves in Fig. 4, depart only by a few dB from the corresponding P-XTC (i.e., $\beta = 0$) filter response (which is a flat curve at 0 dB). More precisely and generally, the maxima and minima of the $E^{[\beta]}_{\text{si}_\|}(\omega)$ spectrum are given by:

$$\begin{aligned}
E^{[\beta]^{\uparrow}}_{\mathrm{si}_{\parallel}} &= \frac{g^2+1}{g^2+\beta+1} \text{ at } \omega\tau_c = n\frac{\pi}{2}, \text{ with } n = 1,3,5,\ldots \\
E^{[\beta]^{\downarrow}}_{\mathrm{si}_{\parallel}} &= \frac{g^4+(\beta-2)g^2+\beta+1}{g^4+2(\beta-1)g^2+(\beta+1)^2} \\
&\qquad \text{at } \omega\tau_c = n\pi, \text{ with } n = 0,1,2,3,4,\ldots
\end{aligned}$$

For the typical ($g = .985$) example shown in the figure, we have, for $\beta = .05$, $E^{[\beta]^{\uparrow}}_{\mathrm{si}_{\parallel}} = -.2$ dB and $E^{[\beta]^{\downarrow}}_{\mathrm{si}_{\parallel}} = -6.1$ dB, showing that even relatively aggressive regularization results in a spectral coloration at the ears that is quite modest compared to the spectral coloration the perfect XTC filter imposes at the loudspeakers.

In sum, we conclude that, while constant-parameter regularization is effective at reducing the amplitude of peaks (including the "low-frequency boost") in the envelope spectrum at the loudspeakers, it typically results in undesirable narrow-band artifacts at higher frequencies and a rolloff of the lower frequencies at the loudspeakers. This non-optimal behavior can be avoided if the regularization parameter is allowed to be a function of the frequency, as we shall see in Section IV.

Before we do so, it is insightful to consider the effects of constant-parameter regularization on the time-domain response of XTC filters.

B. Impulse Response

We start by making the substitution $z = e^{i2\omega\tau_c}$ in Eqs. (25) and (26) to get

$$\begin{aligned}
H^{[\beta]}_{LL}(z) &= H^{[\beta]}_{RR}(z) \\
&= \frac{z^2g^2 - z(\beta+1)}{z^2g^2+g^2-z\left[(g^2+\beta)^2+2\beta+1\right]}, \quad (33) \\
H^{[\beta]}_{LR}(z) &= H^{[\beta]}_{RL}(z) \\
&= \frac{z\left[gz^{-1/2}-g(g^2+\beta)z^{1/2}\right]}{z^2g^2+g^2-z\left[(g^2+\beta)^2+2\beta+1\right]}. \quad (34)
\end{aligned}$$

The two expressions above have the same quadratic denominator, which can be factored as

$$z^2g^2+g^2-z\left[(g^2+\beta)^2+2\beta+1\right] = g^2(z-a_1)(z-a_2),$$

where

$$a_1 = \frac{a-\sqrt{a^2-4g^4}}{2g^2}, \quad a_2 = \frac{a+\sqrt{a^2-4g^4}}{2g^2}, \tag{35}$$

and

$$a = (g^2+\beta)^2+2\beta+1. \tag{36}$$

We can then re-write Eqs. (33) and (34) as

$$\begin{aligned}
H^{[\beta]}_{LL}(z) &= H^{[\beta]}_{RR}(z) \\
&= \left[z - \frac{(\beta+1)}{g^2}\right] \times \\
&\qquad \left(\frac{1}{1-a_1z^{-1}}\right)\left(\frac{1}{z-a_2}\right), \quad (37) \\
H^{[\beta]}_{LR}(z) &= H^{[\beta]}_{RL}(z) \\
&= \left[\frac{z^{-1/2}-(g^2+\beta)z^{1/2}}{g}\right] \times \\
&\qquad \left(\frac{1}{1-a_1z^{-1}}\right)\left(\frac{1}{z-a_2}\right). \quad (38)
\end{aligned}$$

Since $0 < g < 1$, and $\beta \geq 0$, we see from Eqs. (35) and (36) that $0 \leq a_1 < 1$ and $a_2 > 1$, and therefore $|a_1z^{-1}| < 1$ and $a_2 > |z|$. This allows us to express the terms $1/(1-a_1z^{-1})$ and $1/(z-a_2)$ in the last two equations as two convergent power series (whose convergence insures that we have a stable filter), and thus write the last two equations as

$$\begin{aligned}
H^{[\beta]}_{LL}(z) &= H^{[\beta]}_{RR}(z) \\
&= \left[z - \frac{(\beta+1)}{g^2}\right] \times \\
&\qquad \left(\sum_{m=0}^{\infty} a_1^m z^{-m}\right)\left(\sum_{m=0}^{\infty} -a_2^{-m-1}z^m\right) \quad (39) \\
H^{[\beta]}_{LR}(z) &= H^{[\beta]}_{RL}(z) \\
&= \left[\frac{z^{-1/2}-(g^2+\beta)z^{1/2}}{g}\right] \times \\
&\qquad \left(\sum_{m=0}^{\infty} a_1^m z^{-m}\right)\left(\sum_{n=0}^{\infty} -a_2^{-m-1}z^m\right) \quad (40)
\end{aligned}$$

The filter is now in a form that can be readily transformed into a time-domain filter, $\boldsymbol{h}^{[\beta]}$, represented by

$$\boldsymbol{h}^{[\beta]} = \begin{bmatrix} h^{[\beta]}_{LL}(t) & h^{[\beta]}_{LR}(t) \\ h^{[\beta]}_{RL}(t) & h^{[\beta]}_{RR}(t) \end{bmatrix}. \tag{41}$$

We do so by substituting back $e^{i2\omega\tau_c}$ for z in Eqs. (39) and (40), and taking the inverse Fourier transform (IFT) to get

$$\begin{aligned}
h^{[\beta]}_{LL}(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H^{[\beta]}_{LL}(i\omega)e^{i\omega t}d\omega \\
&= h^{[\beta]}_{RR}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H^{[\beta]}_{RR}(i\omega)e^{i\omega t}d\omega \\
&= \left[\delta(t+2\tau_c) - \frac{\beta+1}{g2}\delta(t)\right] * \psi(t), \quad (42)
\end{aligned}$$

$$\begin{aligned} h_{LR}^{[\beta]}(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H_{LR}^{[\beta]}(i\omega)e^{i\omega t}d\omega \\ &= h_{RL}^{[\beta]}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H_{RL}^{[\beta]}(i\omega)e^{i\omega t}d\omega \\ &= \left[\frac{\delta(\tau_c - t)}{g} - \frac{(g^2+\beta)\delta(t+\tau_c)}{g}\right] * \psi(t), \end{aligned} \quad (43)$$

where the asterisk denotes the convolution operation, and $\psi(t)$ is the IFT of the product of the two series appearing in Eqs. (39) and (40), and is given by the following convolution of two trains of Dirac delta functions:

$$\psi(t) = \sum_{m=0}^{\infty} a_1^m \delta(t - 2m\tau_c) * \sum_{m=0}^{\infty} -a_2^{-m-1}\delta(t + 2m\tau_c), \quad (44)$$

We see that the first train evolves forward in time and the second evolves in reverse time.

The impulse response (IR) represented by Eqs. (42) and (43) is plotted in Fig. 5 for three values of β.

The IR of the perfect XTC filter is shown in the top panel of that figure and consists of two trains of decaying and inter-delayed delta functions of opposite sign. Mathematically, it is the special case of $\beta = 0$, for which Eqs. (37) and (38) simplify to

$$H_{LL}^{[P]}(z) = H_{RR}^{[P]}(z) = \frac{1}{1 - a_1 z^{-1}}, \quad (45)$$

$$H_{LR}^{[P]}(z) = H_{RL}^{[P]}(z) = -\frac{g z^{-1/2}}{1 - a_1 z^{-1}}, \quad (46)$$

from which, through the inverse Fourier transform, we recover the IR of the perfect XTC filter derived in Ref. [20]:

$$h_{LL}^{[P]}(t) = h_{RR}^{[P]}(t) = \sum_{n=0}^{\infty} a_1^n \delta(t - 2n\tau_c) \quad (47)$$

$$\begin{aligned} h_{LR}^{[P]}(t) &= h_{RL}^{[P]}(t) \\ &= -g\delta(t-\tau_c) * \sum_{n=0}^{\infty} a_1^n \delta(t - 2n\tau_c), \end{aligned} \quad (48)$$

where $a_1 = g^2$ (obtained by setting $\beta = 0$ in Eqs. (35) and (36)) is the pole of the filter. We see that the perfect XTC IR starts at $t = 0$ with an amplitude of unity and decays to an amplitude $a_1^n = (l_1/l_2)^{2n}$ after a time $2n\tau_c$.

Its physical significance has been discussed by Kirkeby et. al.[20] who, along with Atal et. al.[9] before, recognized the recursive nature of XTC filters. Briefly, a physical appreciation of the perfect crosstalk cancellation IR can be obtained by considering the hypothetical case of a positive pulse whose duration is much smaller than τ_c, fed into only one of the two inputs of the system, say the left input. From Eq. (9), we see that this pulse, $d_L(t)$, is emitted from the left loudspeaker as a series of positive pulses $d_L(t) * h_{LL}(t)$ (corresponding to the filled circles in the top panel of Fig. 5) and from the right loudspeaker as a series of negative pulses $d_L(t) * h_{RL}(t)$ (corresponding to the empty circles in the same plot). These two series of pulses are delayed by τ_c with respect to each other so that after the first positive pressure pulse arrives at the left ear, then reaches the right ear with a slightly smaller amplitude, it is cancelled there by a negative pressure pulse of equal amplitude (that was emitted a time τ_c earlier by the right loudspeaker), which in turn is cancelled at the left ear by a positive pressure pulse, and so on. The net result is that only the first pulse is heard and only at the left ear, i.e., with no crosstalk.

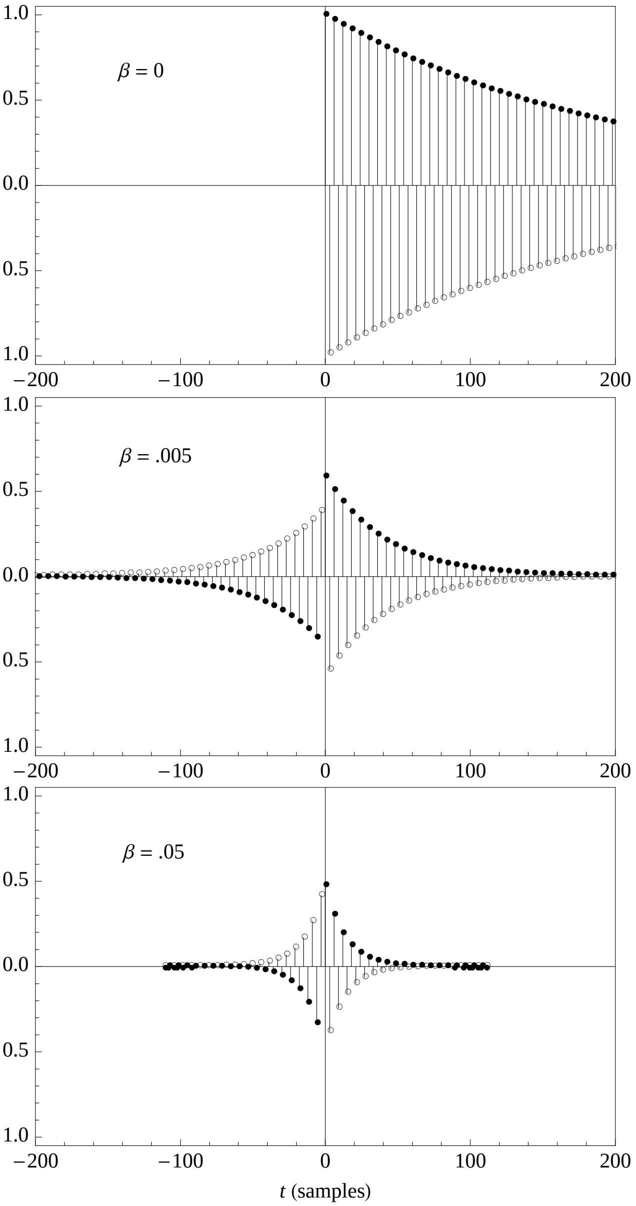

FIG. 5: Impulse responses $h_{LL}^{[\beta]}(t) = h_{RR}^{[\beta]}(t)$ (filled circles) and $h_{LR}^{[\beta]}(t) = h_{RL}^{[\beta]}(t)$ (empty circles) for three values of β. ($g = .985$, $\tau_c = 3$ samples.)

The effects of regularization on the XTC IR can be gleaned from a comparison of the three panels of Fig. 5. When β is finite, the IR has a “pre-echo” part, i.e., it extends in reverse time ($t < 0$) as shown in Fig. 5. As also can be seen in that figure, and inferred from Eq. (44), the delta functions in the $t < 0$ and $t > 0$ parts have opposite signs. With increasing regularization, the $t < 0$ part increases in prominence and the IR becomes shorter in temporal extent, which correspond in the frequency domain to a spectrum with abated peaks.

To insure causality, a time delay must be used to include the $t < 0$ part of the IR. In practice (e.g., when dealing with numerical HRTF inversion), this can be done through a "modelling delay" that accommodates both the non-causal part of the IR and the transmission delay

$$\delta\left(\frac{l_1}{c_s} - t\right)$$

associated with the factor α in Eq. (8).

The length of a filter having a pole close to the unit circle, $|z| = 1$, is inversely proportional to the distance between the pole and the unit circle[38]. As β is increased the poles pull away from the unit circle as per Eqs. (35) and (36), and therefore the length of a finite-β IR is reduced by a factor of

$$\frac{1 - a_1}{1 - g^2}$$

with respect to the length of the perfect XTC IR. This factor (which is based on a_1 since $1 - a_1 < |1 - a_2|$) is accurate as long as $1 - g^2 << 1$ and $1 - a_1 << 1$. For instance, for the IR shown in the middle panel of Fig. 5 we have $\beta = .005$ and $g = .985$, which give $a_1 = .86$ and the IR is about 4.5 times shorter than the perfect XTC IR.

IV. FREQUENCY-DEPENDENT REGULARIZATION

In order to avoid the frequency-domain artifacts discussed in Section III A and illustrated in Fig. 3, we seek an optimization prescription that would cause the envelope spectrum $\hat{S}(\omega)$ to be flat at a desired level Γ (dB) over the frequency bands where the perfect filter's envelope spectrum exceeds Γ (dB). Outside these bands (i.e., below that level), we apply no regularization. This can be stated symbolically as:

$$\hat{S}(\omega) = \gamma \quad \text{if} \quad \hat{S}^{[P]}(\omega) \geq \gamma, \tag{49}$$

$$\hat{S}(\omega) = \hat{S}^{[P]}(\omega) \quad \text{if} \quad \hat{S}^{[P]}(\omega) < \gamma, \tag{50}$$

where the P-XTC envelope spectrum, $\hat{S}^{[P]}(\omega)$, is given by Eq. (16), and

$$\gamma = 10^{\Gamma/20}, \tag{51}$$

with Γ given in dB. We will take $\Gamma \geq 0$ dB and, since Γ cannot exceed the magnitude of the peaks in the $\hat{S}^{[P]}(\omega)$ spectrum, γ is bounded by the inequalities:

$$1 \leq \gamma \leq \frac{1}{1-g}, \tag{52}$$

where the last term is $\hat{S}^{[P]^\uparrow}$, given by Eq. (18).

The frequency-dependent regularization parameter needed to effect the spectral flattening required by Eq. (49) is obtained by setting $\hat{S}^{[\beta]}(\omega)$, given by Eq. (27), equal to γ and solving for $\beta(\omega)$, which is now a function of frequency. Since the regularized spectral envelope, $\hat{S}^{[\beta]}(\omega)$, (which is also $||H^{[\beta]}||$, the 2-norm of the regularized XTC filter) is the maximum of two functions, we get two solutions for $\beta(\omega)$:

$$\beta_{\text{I}}(\omega) = -g^2 + 2g\cos(\omega\tau_c) + \frac{\sqrt{g^2 - 2g\cos(\omega\tau_c) + 1}}{\gamma} - 1, \tag{53}$$

$$\beta_{\text{II}}(\omega) = -g^2 - 2g\cos(\omega\tau_c) + \frac{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}{\gamma} - 1. \tag{54}$$

The first solution, $\beta_{\text{I}}(\omega)$, applies for frequency bands where the out-of-phase response of the perfect filter (i.e., the second singular value, which is the second argument of the max function in Eq. (16)) dominates over the in-phase response (i.e., the first argument of that function):

$$\begin{aligned} S_o^{[P]} &= \frac{1}{\sqrt{g^2 - 2g\cos(\omega\tau_c) + 1}} \\ &\geq S_i^{[P]} = \frac{1}{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}. \end{aligned} \tag{55}$$

Similarly, regularization with $\beta_{\text{II}}(\omega)$ applies for frequency bands where $S_i^{[P]} \geq S_o^{[P]}$. Therefore, we must distinguish between three branches of the optimized solution: two regularized branches corresponding to $\beta = \beta_{\text{I}}(\omega)$ and $\beta = \beta_{\text{II}}(\omega)$, and one non-regularized (perfect-filter) branch corresponding to $\beta = 0$. We call these Branch I, II and P, respectively, and sum up the conditions associated with each as follows:

Branch I: applies where $\hat{S}^{[P]}(\omega) \geq \gamma$ and $S_o^{[P]} \geq S_i^{[P]}$,
and requires setting $\hat{S}(\omega) = \gamma, \quad \beta = \beta_{\text{I}}(\omega)$;

Branch II: applies where $\hat{S}^{[P]}(\omega) \geq \gamma$ and $S_i^{[P]} \geq S_o^{[P]}$,
and requires setting $\hat{S}(\omega) = \gamma, \quad \beta = \beta_{\text{II}}(\omega)$;

Branch P: applies where $\hat{S}^{[P]}(\omega) < \gamma$,
and requires setting $\hat{S}(\omega) = \hat{S}^{[P]}(\omega), \ \beta = 0$.

Following this three-branch division, the envelope spectrum at the loudspeakers, $\hat{S}(\omega)$, for the case of frequency-dependent regularization is plotted as the thick black curve in Fig. 6 for $\Gamma = 7$ dB. This value was chosen because it corresponds to the magnitude of the (doublet) peaks in the $\beta = .05$ spectrum (i.e., $\Gamma = 20\log_{10}(1/2\sqrt{\beta})$), which is also plotted (light solid curve) as a reference for the corresponding case of constant-parameter regularization. (We call a spectrum obtained with frequency-dependent regularization and one obtained with constant-β regularization "corresponding spectra," if the peaks in $\hat{S}^{[\beta]}(\omega)$, whether singlets or doublets, are equal to γ.) It is clear from that figure that the low-frequency boost and the high-frequency peaks of the perfect XTC spectrum, which would be transformed into a low-frequency roll-off and narrow-band artifacts, respectively, by constant-β regularization, are now flat at the desired maximum coloration level, Γ. The rest of the spectrum, i.e., the frequency bands with amplitude below Γ, is allowed to benefit from the infinite XTC level of the perfect XTC filter and the robustness associated with relatively low condition numbers.

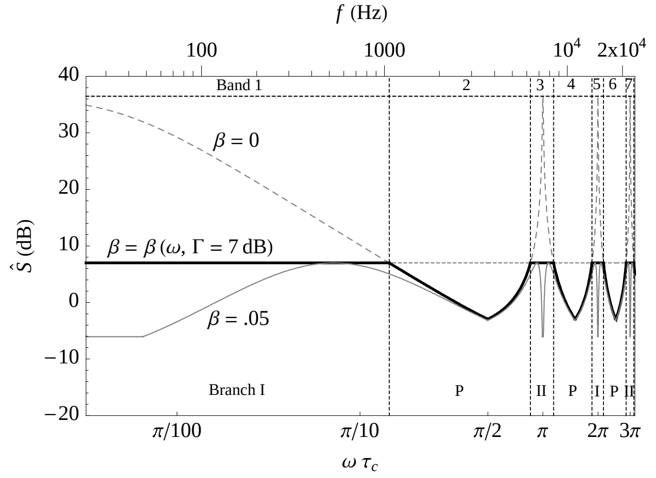

FIG. 6: Envelope spectrum at the loudspeakers, $\hat{S}(\omega)$, for the case of frequency-dependent regularization with $\Gamma = 7$ dB (thick black curve) and for the corresponding reference case of $\beta = .05$ (grey curve). The benchmark case of the perfect XTC filter is also shown (dashed grey curve). The vertical dotted lines show the frequency bounds of the resulting seven bands, which are numbered consecutively at the top of the plot, and labeled with the corresponding branch name at the bottom. (Other parameters are the same as for Fig. 2.)

A. Band Hierarchy

The three-branch prescription therefore splits the audio spectrum into a series of adjacent frequency bands, which we number consecutively starting with Band 1 for the lowest-frequency band. The frequency bounds for each band can be found by setting $\hat{S}^{[P]}(\omega)$, given by Eq. (16), to γ and solving for $\omega\tau_c$. This results in the following hierarchy of bands and their associated frequency bounds:

- Bands $1, 5, 9, 13, 17, \ldots, 4n+1$ belong to Branch I, and are bounded by

$$2n\pi - \phi \leq \omega\tau_c \leq 2n\pi + \phi; \tag{56}$$

- Bands $2, 6, 10, 14, 18, \ldots, 4n+2$ belong to Branch P, and are bounded by

$$2n\pi + \phi \leq \omega\tau_c \leq (2n+1)\pi - \phi; \tag{57}$$

- Bands $3, 7, 11, 15, 19, \ldots, 4n+3$ belong to Branch II, and are bounded by

$$(2n+1)\pi - \phi \leq \omega\tau_c \leq (2n+1)\pi + \phi; \tag{58}$$

- Bands $4, 8, 12, 16, 20, \ldots, 4n+4$ belong to Branch P, and are bounded by

$$(2n+1)\pi + \phi \leq \omega\tau_c \leq 2(2n+1)\pi - \phi; \tag{59}$$

where $n = 0, 1, 2, 3, 4, \ldots$ and

$$\phi = \cos^{-1}\left(\frac{g^2\gamma^2 + \gamma^2 - 1}{2g\gamma^2}\right). \tag{60}$$

For instance, applying this hierarchy to the case of $g = .985$, and $\Gamma = 7$ dB (i.e., $\gamma = 10^{7/20} = 2.24$), shown in Fig. 6, we have the following set of eight consecutive frequency bounds for the seven consecutive bands between $\omega\tau_c = 0$ and 3π: $\{0, 0.45, 2.69, 3.59, 5.83, 6.74, 8.97, 9.42\}$, which correspond to dimensional frequencies, f (Hz) (with $\tau_s = 3$ samples at 44.1 kHz) given by the set: $\{0, 1061.5, 6288.5, 8411.5, 13638.5, 15761.5, 20988.5, 22000\}$, as marked by the vertical lines in Fig. 6. Bands 1 and 5 belong to Branch I and are regularized with $\beta = \beta_{\mathrm{I}}(\omega)$; Bands 3 and 7 belong to Branch II and are regularized with $\beta = \beta_{\mathrm{II}}(\omega)$; and Bands 2, 4, and 6 belong to Branch P and are not regularized. In general, successive bands, starting from the lowest-frequency one, are mapped to the following succession of branches: I, P, II, P, I, P, II, P, ...

B. Frequency Response

The amplitude envelope of the frequency response at the loudspeakers, given by Eqs. (49) and (50), was already shown in Fig. 6. The other optimized metric spectra can be derived as follows:

$$\begin{aligned}
Y^{[O]}_{\mathrm{I}}(\omega) &= Y^{[\beta_{\mathrm{I}}(\omega)]}(\omega), \text{ for Branch-I bands;} && (61)\\
Y^{[O]}_{\mathrm{II}}(\omega) &= Y^{[\beta_{\mathrm{II}}(\omega)]}(\omega), \text{ for Branch-II bands;} && (62)\\
Y^{[O]}_{\mathrm{P}}(\omega) &= Y^{[P]}(\omega), \text{ for Branch-P bands;} && (63)
\end{aligned}$$

where Y represents any of the eight metric spectra we defined in Section II B, the superscript $[O]$ denotes the sought optimized version of that metric spectrum, the subscript I, II, or P denotes one of the three branches, and the superscripts $[\beta_{\mathrm{I}}(\omega)]$ and $[\beta_{\mathrm{II}}(\omega)]$ denote regularization following the formulas for the regularized metric spectra in Section III A, but with β taken to be frequency-dependent according to Eqs. (53) and (54).

For example, following the above hierarchical prescription, and using Eqs. (28), (53), (54), and (17), the optimized crosstalk cancellation spectrum becomes

$$\chi^{[O]}_{\rm I,II}(\omega) \;=\; \mp\frac{\mp\gamma x^2+(g^2+1)\left(x\gamma\pm\sqrt{g^2\mp x+1}\right)}{|x|\left(\gamma g^2\mp\gamma x+\gamma-\sqrt{g^2\mp x+1}\right)}, \tag{64}$$

$$\chi^{[O]}_{\rm P}(\omega) \;=\; \chi^{[p]}(\omega)=\infty, \tag{65}$$

where, for compactness, we have used the definition $x\equiv 2g\cos(\omega\tau_c)$ and combined both branches into one expression using the double subscripts "I,II" and the double sign ($\pm$ or $\mp$) with the top and bottom signs associated with Branches I and II, respectively. Similarly, the optimized version of the ipsilateral frequency response at the ear for a side image, $E_{{\rm si}_\parallel}(\omega)$, becomes

$$E^{[O]}_{{\rm si}_{\parallel\,\rm I,II}}(\omega) = \frac{\pm x\gamma^2(g^2\mp x+1)+(\gamma^2 g+\gamma)\sqrt{g^2\mp x+1}}{g^2\mp x\pm 2\gamma x\sqrt{g^2\mp x+1}+1} \tag{66}$$

$$E^{[O]}_{{\rm si}_{\parallel\rm P}}(\omega)=E^{[P]}_{{\rm si}_\parallel}(\omega)=1. \tag{67}$$

These spectra are plotted in Fig. 7 where it is immediately clear from the $\chi(\omega)$ curves that frequency-dependent regularization yields a significant enhancement of XTC level over that obtained with constant-β regularization. We can also deduce from this plot that the higher the desired minimum level of XTC, the larger is the XTC enhancement over that attained with the corresponding constant-β regularization.

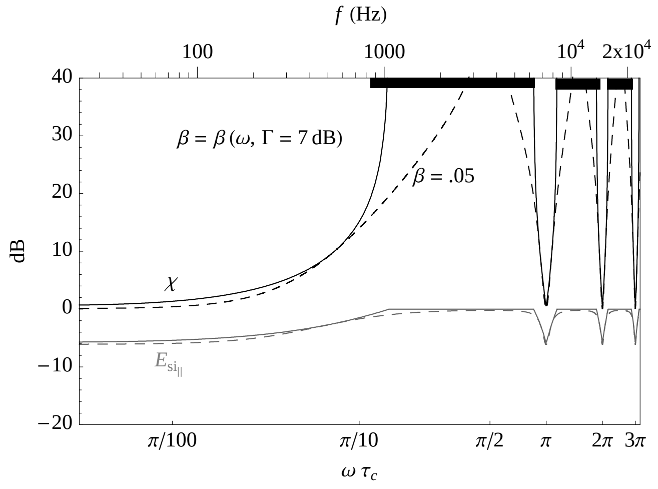

FIG. 7: Crosstalk cancellation spectrum, $\chi(\omega)$ (black curves), and ipsilateral frequency response at the ear for a side image, $E_{{\rm si}_\parallel}(\omega)$ (light curves), for the cases of frequency-dependent regularization (solid curves) and $\beta=.05$ (dashed curves). The frequency ranges for which an XTC level of 20 dB or higher is reached are marked on the top axis by black horizontal bars for the case of $\beta=\beta(\omega)$ with $\Gamma=7$ dB. (Other parameters are the same as for Fig. 2.)

Furthermore, this XTC enhancement occurs with no relative penalty to the frequency response at the ears, as can be seen by comparing the $E_{{\rm si}_\parallel}(\omega)$ spectrum with frequency-dependedent regularization (solid grey curve) to that with $\beta=.05$ (dashed grey curve) in the same figure.

It can be verified through Eqs. (28) and (64) that constant-β regularization yields an XTC level that is equal to that obtained with the corresponding frequency-dependent regularization only at the discrete frequencies at which to the peaks in the corresponding $\hat{S}^{[\beta]}(\omega)$ spectrum are located, i.e., at

$$\begin{aligned}\omega\tau_c \;&=\; n\pi,\ \text{ if }\ \frac{1}{4\gamma^2}<(g-1)^2;\\ &=\; n\pi\,\pm\,\cos^{-1}\left(\frac{g^2-\beta+1}{2g}\right),\\ &\qquad \text{if }\ (g-1)^2\le\frac{1}{4\gamma^2}\ll 1,\\ &\qquad \text{with } n=0,1,2,3,4,\ldots\end{aligned} \tag{68}$$

(where the inequalities are those conditioning singlet or doublet peaks in the corresponding $\hat{S}^{[\beta]}(\omega)$ spectrum, and are derived from Eqs. (29), (30) and (32)). At all other frequencies, frequency-dependent regularization yields superior XTC performance to that obtained with constant-β regularization. This behavior, which can also be seen graphically in the $\chi(\omega)$ curves of Fig. 7, is due to the fact that forcing the envelope spectrum to be flat (in bands belonging to Branches I and II) through frequency-dependent regularization clamps the effort penalty term in the cost function (second term in the sum in Eq. (23)) leading to a minimization of the performance error. This in turn leads to a maximization of XTC level, which exceeds the corresponding constant-β XTC level at all frequencies (except at those given by Eq. (68), where both corresponding $\hat{S}$ spectra reach the same value, γ), since the corresponding constant-β envelope, $\hat{S}^{[\beta]}(\omega)$, is lower than (or equal to) γ (as seen in Fig. 6).

Therefore, we conclude that if we define XTC filter optimization as "the maximization of XTC performance for a desired tolerable level of spectral coloration" as we did earlier, only frequency-dependent regularization leads to an optimal XTC filter over all frequencies, while constant-β regularization leads to an XTC filter that is optimized only at the discrete frequencies given by Eq. (68).

C. Impulse Response: The Band-Assembled Crosstalk Cancellation Hierarchy Filter

In the frequency domain, the optimized XTC filter is given by the following matrix:

$$\boldsymbol{H}^{[O]} = \begin{bmatrix} H^{[O]}_{LL}(i\omega) & H^{[O]}_{LR}(i\omega) \\ H^{[O]}_{RL}(i\omega) & H^{[O]}_{RR}(i\omega) \end{bmatrix}, \tag{69}$$

whose elements are derived following the same hierarchical prescription (i.e., Eqs. (61)-(63)) we used to get the optimized metric spectra, namely by substituting $\beta_1(\omega)$ and $\beta_1(\omega)$ from Eqs. (53) and (54), and $\beta = 0$ into each of Eqs. (25) and (26), to get the Branch I, Branch II and Branch P versions of the filter's matrix elements. This leads to

$$\begin{aligned} &H^{[O]}_{LL_{\rm I,II}}(i\omega) = H^{[O]}_{RR_{\rm I,II}}(i\omega) \\ &\quad = \frac{\gamma^2\left[\pm x - g^2\left(1+e^{2i\omega\tau_c}\right)\right] + \gamma\sqrt{g^2 \mp x + 1}}{g^2 \pm x\left(2\gamma\sqrt{g^2 \mp x+1}-1\right)+1}, \end{aligned} \tag{70}$$

$$\begin{aligned} &H^{[O]}_{LR_{\rm I,II}}(i\omega) = H^{[O]}_{RL_{\rm I,II}}(i\omega) \\ &\quad = \frac{\mp\gamma^2\left[\pm x - g^2\left(1+e^{2i\omega\tau_c}\right)\right] + g\gamma e^{i\omega\tau_c}\sqrt{g^2 \mp x + 1}}{g^2 \pm x\left(2\gamma\sqrt{g^2 \mp x+1}-1\right)+1}, \end{aligned} \tag{71}$$

$$H^{[O]}_{LL_{\rm P}}(i\omega) = H^{[O]}_{LL_{\rm P}}(i\omega) = H^{[P]}_{LL}(i\omega) = H^{[P]}_{RR}(i\omega), \tag{72}$$

$$H^{[O]}_{LR_{\rm P}}(i\omega) = H^{[O]}_{RL_{\rm P}}(i\omega) = H^{[P]}_{LR}(i\omega) = H^{[P]}_{RL}(i\omega), \tag{73}$$

where, again,

$$x \equiv 2g\cos(\omega\tau_c),$$

and we have followed the same subscript and sign conventions used to compact the XTC spectrum in Eq. (64). Eqs. (72) and (73) give the Branch-P elements of the matrix of the optimized filter, which are also the elements of the perfect filter's matrix given by Eqs. (45) and (46), whose inverse Fourier transforms had given us the IRs expressed in Eqs. (47) and (48). Therefore we need to derive only the IRs associated with Branches I and II of the optimized filter.

To do so, we follow, albeit through more cumbersome algebra, the same approach we used to obtain the constant-β IRs in Section III B; namely, we seek to factor the frequency-domain response of the filter into a product of terms, whose IFT can be readily found, or which can be expressed as convergent series of functions whose IFT can be readily found. The complete IR is then the convolution of the IFTs of all the terms in the factored frequency-domain response of the filter. The challenge is to carry out the factorization in such a way that all the invoked power series expansions converge over the parameter space of interest.

The derivation is carried out in Appendix A, where we also discuss the convergence of the adopted series expansions. The resulting filter in the time domain is given by the following two IRs:

$$h^{[O]}_{LL_{\rm I,II}}(t) = h^{[O]}_{RR_{\rm I,II}}(t) = (\psi_0 + \gamma\psi_1) * \psi_a, \tag{74}$$

$$h^{[O]}_{LR_{\rm I,II}}(t) = h^{[O]}_{RL_{\rm I,II}}(t) = [\mp\psi_0 + \gamma g\delta(t+\tau_c) * \psi_1] * \psi_a, \tag{75}$$

where

$$\psi_a = \pm\psi_2 * \psi_3 \pm (\psi_1 \mp \psi_4) * \psi_5\psi_6(c_1) * \psi_6(c_2), \tag{76}$$

$$\begin{aligned}\psi_0 = {} & -g^2\gamma^2\delta(t) \pm g\gamma^2\delta(\tau_c - t) \pm g\gamma^2\delta(t+\tau_c) \\ & -g^2\gamma^2\delta(t+2\tau_c),\end{aligned} \tag{77}$$

$$\begin{aligned}\psi_1 = {} & \sum_{m=0}^{\infty}\begin{pmatrix}\frac{1}{2}\\ m\end{pmatrix}(\mp g)^m\left(g^2+1\right)^{\frac{1}{2}-m} \times \\ & \sum_{k=0}^{m}\begin{pmatrix}m\\ k\end{pmatrix}\delta(2k\tau_c - t - m\tau_c),\end{aligned} \tag{78}$$

$$\begin{aligned}\psi_2 = {} & \pm\frac{1}{4g\gamma}\sum_{m=0}^{\infty}\begin{pmatrix}-\frac{1}{2}\\ m\end{pmatrix}(-1)^m \times \\ & \sum_{k=0}^{2m}\begin{pmatrix}2m\\ k\end{pmatrix}(-1)^{k+m}4^{-m}\delta(t+2k\tau_c-2m\tau_c),\end{aligned} \tag{79}$$

$$\begin{aligned}\psi_3 = {} & \sum_{m=0}^{\infty}\begin{pmatrix}-\frac{1}{2}\\ m\end{pmatrix}(\mp g)^m\left(g^2+1\right)^{-\frac{1}{2}-m} \times \\ & \sum_{k=0}^{m}\begin{pmatrix}m\\ k\end{pmatrix}\delta(2k\tau_c - t - m\tau_c),\end{aligned} \tag{80}$$

$$\psi_4 = 2g\gamma\delta(\tau_c - t) + 2g\gamma\delta(t+\tau_c), \tag{81}$$

$$\begin{aligned}\psi_5 = {} & \pm\frac{1}{(4g\gamma)^3}\sum_{m=0}^{\infty}\begin{pmatrix}-\frac{3}{2}\\ m\end{pmatrix}(-1)^m \times \\ & \sum_{k=0}^{2m}\begin{pmatrix}2m\\ k\end{pmatrix}(-1)^{k+m}4^{-m}\delta(t+2k\tau_c-2m\tau_c),\end{aligned} \tag{82}$$

$$\begin{aligned}\psi_6(c) = {} & \sum_{p=0}^{\infty}\left(\frac{\pm c}{2g}\right)^p\sum_{n=0}^{\infty}\begin{pmatrix}-\frac{p}{2}\\ n\end{pmatrix}(-1)^n \times \\ & \sum_{k=0}^{2n}\begin{pmatrix}2m\\ k\end{pmatrix}(-1)^{k+m}4^{-m}\delta(t+2k\tau_c-2m\tau_c).\end{aligned} \tag{83}$$

with the constants c_1 and c_2 given by

$$c_1 = \frac{\sqrt{16\gamma^2(g^2+1)+1} \mp 1}{8\gamma^2}, \tag{84}$$

$$c_2 = \frac{-\sqrt{16\gamma^2(g^2+1)+1} \mp 1}{8\gamma^2}. \tag{85}$$

The impulse responses are valid for values of γ and g that satisfy the condition:

$$\max\left(\frac{\sqrt{5+\sqrt{5}}}{2\sqrt{g^2+1}},1\right) \le \gamma \le \frac{1}{1-g}, \tag{86}$$

which is shown graphically as a region plot in Fig. 9, in Appendix A.

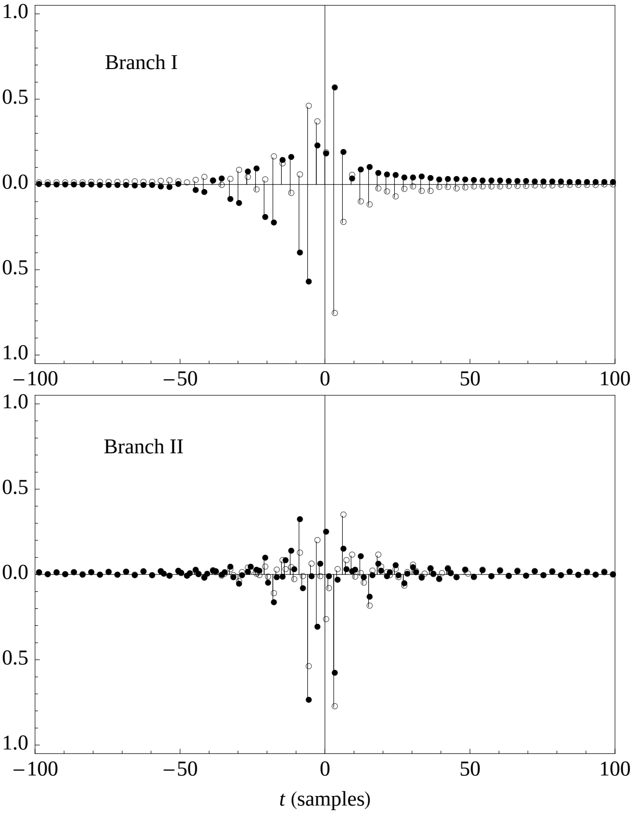

FIG. 8: Impulse response of the optimal XTC filter: $h^{[O]}_{LL}(t) = h^{[O]}_{RR}(t)$ (filled circles) and $h^{[O]}_{LR}(t) = h^{[O]}_{RL}(t)$ (empty circles), for Branch I (top panel) and Branch II (bottom panel). ($\Gamma = 7$ dB, $g = .985$, and $\tau_c = 3$ samples, as in Fig. 2.)

The impulse responses for Branch-I and Branch-II of this optimal filter are shown in Fig. 8 for our typical case of $g = .985$ and $\tau_c = 3$ samples, and, along with the perfect filter IR shown in the top panel of Fig. 5, completely specify the optimal XTC filter.

Compared to the corresponding ($\beta = .05$) finite-beta IR in the bottom panel of Fig. 5, the optimal XTC IRs shown in Fig. 8 are more complex in their structure. Furthermore, each IR consists of a train of deltas that are spaced by τ_c as opposed to the $2\tau_c$ intervals we had for the perfect and finite-beta filters.

These IRs are difficult to interpret physically because, as they stand, they also include the time response associated, in the frequency domain, with frequency bands where the IR is not valid. This is illustrated in Appendix B, in the bottom panel of Fig. 10, where the envelope spectrum obtained from the Fourier transform of the Branch-I optimal IR is compared to the expected flat envelope spectrum, $\hat{S}^{[O]}_I(\omega) = \gamma$. The agreement is excellent only in the bands belonging to the branch for which the IR is intended (which, in the case illustrated in that plot, are the first and fifth bands). In other bands, not only is the IR not valid, but, as discussed in the appendices, its application may lead to singularities associated with the divergence of some of the series that constitute it (see for instance the singularities appearing in the Branch-P bands in Fig. 10).

Therefore, in principle, the application of the optimal filter requires that, prior to XTC filtering, the recorded signal, $[d_{L_i}(t), d_{R_i}(t)]^T$, be passed through a crossover filter whose crossover frequencies are set to the band bounds given by the hierarchical prescription in Eqs. (56)-(59). The resulting bands are then assembled into three groups (I, II and P) according to their branch identity. The combined recorded stereo signals in each group can thus be represented by a vector $[d_{L_i}(t), d_{R_i}(t)]^T$, where the index i stands for Branch I, II or P. The loudspeakers source vector, in the time domain, needed for optimal crosstalk cancellation is then given by the time-domain version of Eq. (9):

$$\begin{bmatrix} v_L(t) \\ v_R(t) \end{bmatrix} = \sum_i \left(\begin{bmatrix} h^{[O]}_{LL_i}(t) & h^{[O]}_{LR_i}(t) \\ h^{[O]}_{RL_i}(t) & h^{[O]}_{RR_i}(t) \end{bmatrix} * \begin{bmatrix} d_{L_i}(t) \\ d_{R_i}(t) \end{bmatrix} \right), \tag{87}$$

where the summation is over the three branches, and the convolution operates in the same fashion as matrix multiplication.

Causality is insured by calculating the IRs with a "pre-delay," starting back at a time $t < 0$, whose exact temporal extent is not important as long as it allows the inclusion of the salient part of the IR. For the IRs in Fig. 8, this pre-delay should start at about $t = -100$ samples.

V. APPLICATION AND PRACTICAL CONSIDERATIONS

While the first goal of the preceding analyses was to provide insight into the theory and fundamentals of XTC optimization, the resulting optimized IR can offer practical benefits in audio applications where the spatial fidelity of a recording is degraded by the unintended crosstalk inherent in playback through loudspeakers. As we argued in Section I A, this is the case of not only pure binaural recordings made with dummy head microphones, but also of the vast majority of standard stereo recordings, since they generally contain ILD and ITD cues, which would be degraded by unintended crosstalk.

A. The Value of Analytical XTC Filters

Analytical XTC filters cannot rival the performance of numerical HRTF-based XTC filters in ideal situations, i.e., when 1) sound reflections in the listening room are negligible or non-existent (anechoic or semi-anechoic environments), 2) the recording was made with the individualized inverted HRTF of the listener, 3) the XTC filter includes the individualized inverted HRTF of the listener, and 4) the listener's head is constrained in a restricted sweet spot. Any departure from these idealities would cause the effective XTC level at the listener's ears to drop, and the spectral coloration that is necessarily imposed at the loudspeakers to become more audible at the listener's ears.

Since in many, if not most, practical listening situations in non-anechoic environments all of the four idealities listed above are compromised to a certain degree, the practically achievable XTC level of numerical HRTF-based XTC filters seldom exceeds 13 dB over a wide frequency range[3]. An optimal analytical XTC filter, even one based on a free-field model, such as the one derived in the previous section, can become competitive especially in situations where it is calculated for, and used with, a loudspeaker span that is small enough to diminish the relative importance of head-shadowing effects. In such applications, an optimal analytical XTC filter can offer the following advantages over a numerical HRTF-based XTC filter:

1. The simplicity of using a single filter for all individuals.
2. Shorter filters which incur lower CPU loads on the digital processor.
3. Low spectral coloration for listeners inside and outside the sweet spot (and the associated decrease in the physical loading of the transducers).

(The third advantage could, in principle, be neutralized by applying the optimal regularization method, described in the previous section, to the design of a numerical HRTF-based XTC – at the price of eroding some of the advantages associated with the use of an individualized HRTF.)

With this justification for the usefulness of analytically-derived optimal XTC filters, we turn our attention to some practical issues related to their specific design and their application to real listening situations.

B. Filter Design Strategy

Of course, filter design strategies depend on performance requirements (desired maximum tolerable coloration level or minimum XTC level) and the specifics and constraints of the listening configuration (constraints on the listening distance, l, and the loudspeaker span, Θ, and to some extent, the sound reflection characteristics of the listening room).

One approach to filter design is to start with the specification of the maximum tolerable coloration level, that is, Γ in dB. For instance, for critical (e.g., audiophile) listening and audio mastering applications, it may not be desirable to have Γ exceed 3-5 dB, while for home-theatre applications, audio (spectral) fidelity may be intentionally compromised with higher values of Γ for the advantage of having more XTC headroom for reproducing surround effects with the two loudspeakers.

The choice of loudspeaker span is particularly important. In cases where it is constrained to a set value, as for compatibility with the so-called "standard stereo triangle", i.e., $\Theta = 60°$, the value of Θ becomes a fixed input to the design process and is used, along with l, to calculate g and τ_c from Eqs. (3)-(6). (The inequality in Eq. (86), which is typically easy to satisfy, must hold for that particular combination of $\gamma = 10^{\Gamma/20}$ and g. If not, one of the input parameters, usually Γ, must be adjusted accordingly before proceeding further with the design). In cases where Θ is not constrained to a preset value, it becomes a useful variable in the filter design process and can be used to simplify the filter, as discussed in Section V C below.

With γ, g, and τ_c specified, one has all the parameters needed to calculate the spectra associated with the XTC optimal filter, as described in Sections IV A and IV B, and thus evaluate the various aspects of the filter. (These evaluations are more conveniently done in terms of the dimensional frequency, f, in Hz, by selecting the intended sampling rate.) In particular, a plot of the XTC spectrum according to Eqs. (64) and (65) allows the evaluation of the XTC performance of the filter (defined as the frequency extent over which a desired minimum XTC level is reached or exceeded) which, by virtue of the implicit optimization (i.e., minimization of the cost function in Eq. (23)), is the maximum achievable XTC performance for that particular set of input parameters. If the calculated XTC performance is judged by some empirical standards to be above that achievable in the intended listening environment (for instance, sound reflections in a reverberant room may limit the achievable XTC to only a few dB over a good part of the audio spectrum), the calculation can be repeated with a lower value of Γ, thus leading to even higher spectral fidelity. Conversely, a lower than desired XTC performance can be amended by raising Γ.

Once the target XTC performance and coloration level are reached, one proceeds to the time domain by calculating the Branch-P IRs from Eqs. (42)-(44), and the Branch-I, and II IRs from Eqs. (74)-(85). The loudspeakers source vector can then be calculated according to Eq. (87), following the prescription given in the text preceding that equation, i.e., by appropriately convolving the 3-part IR with the recorded stereo signal after having passed the latter through a crossover filter whose crossover frequencies are set to the band bounds given in Eqs. (56)-(59). The convolution operations can be carried out digitally, and in real time if desired, using a digital convolution plugin. (Such software plugins rely on FFT-based algorithms[39, 40] for fast convolution and have become readily available in the commercial and public domains for use as IR-based reverb processors.)

C. Simplified Implementation

An XTC system consisting of the properly configured crossover filter, the three XTC IR matrices, and the multiple instances of convolution plugins, can be considered as a single filter, having stereo inputs and outputs, which acts as a linear operator. Therefore, once assembled, it can be "rung" once by a single delta impulse, applied to one of its two inputs, and the recorded stereo output would then represent one of the two columns of the 2x2 IR matrix of the entire filter. Because of symmetry, the other column of the IR matrix is obtained by simply flipping the two recorded outputs. This results in a single IR, representing the entire three-branch multi-band filter, and simplifies any future application of Eq. (87) to a simpler one (with no crossover filtering) in which the summation and indices are foregone.

D. The Role of Loudspeaker Span

Another important simplification arises in applications where the loudspeaker span, $\Theta = 2\theta$, is not constrained to a preset value, such at the 60° of the standard stereo triangle, and therefore can be a variable in the filter design process. Since τ_c depends on the loudspeaker span, the bounds of the bands can be moved by varying θ. By setting θ equal to a particular value, θ^*, the higher bound of the second band (which belongs to Branch P) can be made to coincide with a cutoff frequency, f_c, above which XTC is psychoacoustically not needed. Such a band-limited optimal XTC filter has the advantage that it requires only a 2-band crossover filter, and its IR consists of only the Branch-I and Branch-P parts, thus leading to significant simplifications in the design and implementation of the filter.

To find an expression for θ^* as a function of f_c, under the typically valid approximations $g \simeq 1$ and $l \gg \Delta r$, we set $\omega\tau_c$ equal to the upper bound of the second band (which, from Eq. (57), is $\pi - \phi$), use Eq. (21), and solve for θ, to get

$$\theta^* \simeq \sin^{-1}\left[\frac{c_s\left(\pi - \cos^{-1}\left[\frac{2\gamma^2-1}{2\gamma^2}\right]\right)}{2\pi f_c \Delta r}\right]. \tag{88}$$

A number of studies[27, 41] have suggested that XTC above a frequency of about 6 kHz is not critical or perhaps even necessary. Therefore, we set f_c to that value in the above equation, solve for θ^*, design the filter for a loudspeaker span of $2\theta^*$, use a 2-band crossover filter to separate the first two bands, apply the Branch-I and Branch-P parts of the filter to the first and second bands, respectively, and allow the part of the audio spectrum above f_c to bypass the filter.

It is relevant to mention in the context of loudspeaker span that keeping Θ small offers advantages that have been recognized since Kirkeby and co-workers presented their analysis[20] of the "stereo dipole" configuration, which has a span of only 10°. Objective and subjective evaluations of the effects of loudspeaker span in XTC systems have indicated that such a low-Θ configuration gives a larger sweet spot than that obtained with larger loudspeaker spans[18]. This can be attributed to the relative insensitivity of the path length difference, Δl, to head movements when the span is small. On the other hand, the same study favored larger spans partly because increasing the span, with the distance l fixed, lowers the value of g and consequently decreases the magnitude of the coloration peaks and condition numbers. We should however expect, in light of our study of regularization, that an optimal XTC filter in which regularization is used to flatten these peaks and lower the condition numbers, while maintaining good XTC performance, should tip the balance in favor of lower values of Θ. This remains to be verified experimentally.

Another argument in favor of small loudspeaker spans is particular to the use of analytical filters based on a free-field model, such as those discussed in this paper. Since the free-field model ignores the presence of the listener's head, it should be expected that filters based on it perform better when the effects of head shadowing are minimized. This can be achieved by decreasing the span angle as can be seen, for instance, in Fig. 3.13 of Ref. [27], where the inter-aural transfer function (the ratio of the frequency responses at the two ears) of a typical human head, measured as a function of the azimuthal position of a sound source, is small (about -2dB) and flat (within 2 dB) for a small horizontal source azimuth ($\theta = 5°$), but worsens with increasing azimuths.

E. An Example

To illustrate the above design guidelines and discussions, we give the example of a listening situation whose only two design requirements are a distance $l = 1.6$ m and a maximum coloration level of $\Gamma = 7$ dB. From Eq. (88), with $f_c \simeq 6$ kHz, and $\Delta r = 15$ cm[42], we get $\theta = 9°$, which we take as half the loudspeaker span. From Eqs. (3)-(6), we can then calculate $g = .985$ and $\tau_c = 3$ samples at a sampling rate of 44.1 kHz. These are precisely the dimensional and non-dimensional parameters chosen for the calculations that are illustrated in the plots throughout this paper. The Branch-P and Branch-I IRs are therefore given by those shown in the top panels of Figs. 5 and 8, respectively. The Branch-II IR is not needed as the XTC filter is limited to 6 kHz, which, by design, was made to be the upper bound of the second band (Branch-P). The spectra associated with this filter are given by the solid curves in Figs. 6 and 7, with the dimensional frequency read off the top axes of the plots, up to the cuttoff frequency of 6 kHz. In particular, we note that the XTC performance (top curve in Fig. 7), exceeds 20 dB for a wide range of frequencies that extends from the 6 kHz cuttoff down to 850 Hz, then drops off with decreasing frequency, reaching 5 dB at 290 Hz.

VI. SUMMARY

We distill the main points and findings of this work in the following bullets.

- 3-D reproduction of binaural audio with two loudspeakers requires cancellation of the crosstalk between the loudspeakers and the contralateral ears of the listener. A perfect XTC filter (i.e., one with infinite crosstalk cancellation) can be easily designed but causes severe spectral coloration to the sound emitted by the loudspeakers due to the ill-conditioned inversion of the system's transfer function.

- The coloration produced by the perfect XTC filter consists of peaks in the frequency spectrum that can typically exceed 30 dB, and thus strain the playback transducers and significantly reduce the dynamic range of the playback system. Furthermore, the coloration is heard throughout the listening space and, due to extreme sensitivity to errors in the system, it is also heard by the listener in the sweet spot.

- Using a two-source free-field model, we have shown that constant-parameter regularization, which has been used previously to design HRTF-based XTC systems, can tame these peaks but can produce a bass roll-off and high-frequency artifacts in the filter's frequency response. Furthermore, we demonstrated that constant-beta regularization does not lead to the optimization of XTC filters across all frequencies, but rather only at discrete, widely-spaced frequencies.

- Full optimization can be achieved through frequency-dependent regularization, and requires that the audio spectrum be divided into a hierarchical set of adjacent frequency bands, each of which belonging to one of three solution branches that make up the complete optimal filter. We derived analytical expressions for the three branches of the filter in terms of series expansions, which we showed are convergent for typical listening situations. The corresponding impulse responses were then obtained analytically, and expressed as convolutions of trains of Dirac deltas.

- Aside from seeking fundamental insight into the nature and characteristics of optimal XTC filters, we addressed a number of issues related to their application. In particular, we argued that analytical filters derived under the simplifying assumptions of a free-field model can be useful in practical situations where individualized HRTF-based XTC filters are either too cumbersome to implement or not needed to attain the XTC levels required for enhancing the spatial fidelity of playback in non-anechoic environments. We described a strategy for designing such optimal filters that meets practical design requirements, and we gave an illustrative example for a typical listening configuration.

Acknowledgments

The author wishes to thank Ralph Glasgal for introducing him to the XTC problem, Angelo Farina for his X-Volver convolution plugin, Robin Miller for his independent analytical auditioning of the filters, and J. S. Bach for his Mass in B Minor.

APPENDIX A: DERIVATION OF THE IMPULSE RESPONSE OF THE OPTIMAL XTC FILTER

Here we carry out the derivation of Eqs. (74) to (83) following the approach outlined in Section IV C.

We start by factoring the expressions appearing in Eqs. (70) and (71), which, we note, have the same denominator, into the following product of terms:

$$H^{[O]}_{LL_{\mathrm{I,II}}}(i\omega) = H^{[O]}_{RR_{\mathrm{I,II}}}(i\omega) = \left(\Psi_0 + \gamma\Psi_1\right)\Psi_a \qquad \text{(A1)}$$

$$H^{[O]}_{LR_{\mathrm{I,II}}}(i\omega) = H^{[O]}_{RL_{\mathrm{I,II}}}(i\omega) = \left(\mp\Psi_0 + \gamma g e^{i\omega\tau_c}\Psi_1\right)\Psi_a, \qquad \text{(A2)}$$

where

$$\Psi_0 = \gamma^2\left[\pm x - \left(1 + e^{2i\omega\tau_c}\right)g^2\right], \qquad \text{(A3)}$$

$$\Psi_1 = \sqrt{g^2 \mp x + 1}, \qquad \text{(A4)}$$

$$\Psi_a = \frac{1}{g^2 \pm x\left(2\gamma\sqrt{g^2 \mp x + 1} - 1\right) + 1}. \qquad \text{(A5)}$$

The term Ψ_a can be factored as

$$\Psi_a = \pm\Psi_2\Psi_3 \pm (\Psi_1 \mp \Psi_4)\Psi_5\Psi_6(c_1)\Psi_6(c_2),$$

where

$$\Psi_2 = \frac{1}{2\gamma x}, \qquad \text{(A6)}$$

$$\Psi_3 = \frac{1}{\sqrt{g^2 \mp x + 1}}, \qquad \text{(A7)}$$

$$\Psi_4 = 2\gamma x, \qquad \text{(A8)}$$

$$\Psi_5 = \frac{1}{8\gamma^3 x^3}, \qquad \text{(A9)}$$

$$\Psi_6(c) = \frac{1}{1 - c x^{-1}}, \qquad \text{(A10)}$$

and

$$c_1 = \frac{\sqrt{16\gamma^2(g^2+1)+1} \mp 1}{8\gamma^2}, \tag{A11}$$

$$c_2 = \frac{-\sqrt{16\gamma^2(g^2+1)+1} \mp 1}{8\gamma^2}. \tag{A12}$$

In the time domain, the filter expressed by Eqs. (A1) and (A2) becomes:

$$h^{[O]}_{LL_{\text{I,II}}}(t) = h^{[O]}_{RR_{\text{I,II}}}(t) = (\psi_0 + \gamma\psi_1) * \psi_a, \tag{A13}$$

$$h^{[O]}_{LR_{\text{I,II}}}(t) = h^{[O]}_{RL_{\text{I,II}}}(t) = [\mp\psi_0 + g\gamma\delta(t+\tau_c) * \psi_1] * \psi_a, \tag{A14}$$

where

$$\psi_a = \pm\psi_2 * \psi_3 \pm (\psi_1 \mp \psi_4) * \psi_5\psi_6(c_1) * \psi_6(c_2). \tag{A15}$$

The ψ_i terms are functions of time, and are the IFTs of the Ψ_i terms, which are functions of frequency.

We now seek the IFT of each of the Ψ_i terms given above.

• Ψ_0: The IFT of the expression in Eq. (A3) can be readily found by substituting back $2g\cos(\omega\tau_c)$ for x and carrying out the IFT integration:

$$\begin{aligned}\psi_0(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \gamma^2 \left[\pm 2g\cos(\omega\tau_c) - g^2 - e^{2i\omega\tau_c}\right] d\omega \\ &= -g^2\gamma^2\delta(t) \pm g\gamma^2\delta(\tau_c - t) \pm g\gamma^2\delta(t+\tau_c) \\ &\quad -g^2\gamma^2\delta(t+2\tau_c). \end{aligned} \tag{A16}$$

• Ψ_1: Making the substitution $b = g^2+1$ in Eq. (A4), we get

$$\Psi_1 = \sqrt{b \mp x}, \tag{A17}$$

which can be expressed as the series expansion

$$\Psi_1 = \sum_{m=0}^{\infty} \binom{\frac{1}{2}}{m} (\mp x)^m b^{\frac{1}{2}-m} \tag{A18}$$

where we have used the binomial coefficient

$$\begin{aligned}\binom{k}{m} &= \frac{k!}{m!(k-m)!} \quad \text{if} \quad 0 \le m \le k \\ &= 0 \quad \text{if} \quad m<0 \text{ or } k<m.\end{aligned}$$

Since $0 < g < 1$, we have $x = 2g\cos(\omega\tau_c)| < g^2+1 = b$, and the series in Eq. (A18) always converges. However, as $g \to 1$, $b \to 2$, and when $\omega\tau_c \to n2\pi$ with $n = 0,1,2,3,4,\ldots$, $x \to b$ and the series converges slowly. Replacing x and b by their explicit values, we get

$$\Psi_1 = \sum_{m=0}^{\infty} \binom{\frac{1}{2}}{m} (\mp 2)^m g^m (g^2+1)^{\frac{1}{2}-m} \cos^m(\omega\tau_c). \tag{A19}$$

Since $\cos^m(\omega\tau_c)$ can be written as the finite sum

$$\cos^m(\omega\tau_c) = \sum_{k=0}^{m} \binom{m}{k} 2^{-m} e^{-i(2k-m)\omega\tau_c}, \tag{A20}$$

and since the IFT of $e^{-i(2k-m)\omega\tau_c}$ is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i(2k-m)\omega\tau_c}\, d\omega = \delta(2k\tau_c - t - m\tau_c),$$

the IFT of Ψ_1 can be expressed as

$$\begin{aligned}\psi_1(t) &= \sum_{m=0}^{\infty} \binom{\frac{1}{2}}{m} (\mp g)^m \left(g^2+1\right)^{\frac{1}{2}-m} \times \\ &\qquad \sum_{k=0}^{m} \binom{m}{k} \delta(2k\tau_c - t - m\tau_c). \end{aligned} \tag{A21}$$

• Ψ_2: Explicitly, Eq. (A6) is

$$\Psi_2 = \frac{\sec(\omega\tau_c)}{4g\gamma}.$$

The problem is that the IFT of $\sec(\omega\tau_c)$ cannot be expressed in terms of real delta functions. However, the function $\sec(\omega\tau_c)$ can be expressed as

$$\begin{aligned}\sec(\omega\tau_c) &= \frac{1}{\sqrt{1-\sin^2(\omega\tau_c)}} \\ &\text{if} \quad n2\pi - \frac{\pi}{2} < \omega\tau_c < n2\pi + \frac{\pi}{2} \\ &\text{with} \quad n = 0,1,2,3,4,\ldots\end{aligned} \tag{A22}$$

Furthermore, we note that since

$$1 \le \gamma \le \frac{1}{1-g} \quad \text{and} \quad 0 < g < 1, \tag{A23}$$

the arguments of the inverse cosine function in Eq. (60) obeys the condition:

$$\frac{g^2\gamma^2 + \gamma^2 - 1}{2g\gamma^2} \ge 0 \tag{A24}$$

which leads us to write

$$0 \le \phi \le \frac{\pi}{2}. \tag{A25}$$

In light of this expression and Eq. (56), we conclude that the conditions for the validity of Eq. (A22) are always satisfied in Branch-I bands.

Similarly, we find that $\sec(\omega\tau_c)$ can be expressed as $-1/\sqrt{1-\sin^2(\omega\tau_c)}$ for conditions that are always satisfied for Branch-II bands. Therefore, we can write

$$\sec(\omega\tau_c) = \pm\frac{1}{\sqrt{1-\sin^2(\omega\tau_c)}}, \tag{A26}$$

for which we wish to use the expansion

$$\frac{1}{\sqrt{1-u}} = \sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} (-1)^m u^m. \tag{A27}$$

However, this series converges only for $|u| < 1$. For our particular case, $u = \sin^2(\omega\tau_c)$ and the series diverges at $\omega\tau_c = n\pi/2$, with $n = 1, 3, 5, 7, \ldots$ From the band division conditions in Eqs. (56) and (56) we see that these values of $\omega\tau_c$ are always outside Branch-I and Branch-II bands; therefore the convergence of the series is assured and this allows us to express Eq. (A26) as

$$\sec(\omega\tau_c) = \pm \sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} (-1)^m \sin^{2m}(\omega\tau_c). \tag{A28}$$

Since $\sin^{2m}(\omega\tau_c)$ can be written as the finite sum

$$\sin^{2m}(\omega\tau_c) = \sum_{k=0}^{2m} \binom{2m}{k} (-1)^{k+m} 4^{-m} e^{2i(k-m)\omega\tau_c}, \tag{A29}$$

and since the IFT of $e^{2i(k-m)\omega\tau_c}$ is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{2i(k-m)\omega\tau_c}\, d\omega = \delta(t + 2k\tau_c - 2m\tau_c), \tag{A30}$$

the IFT of Ψ_2 can be expressed as

$$\begin{aligned}\psi_2 \;=\; &\pm\frac{1}{4g\gamma}\sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} (-1)^m \times \\ &\sum_{k=0}^{2m} \binom{2m}{k} (-1)^{k+m} 4^{-m} \delta(t + 2k\tau_c - 2m\tau_c).\end{aligned} \tag{A31}$$

- Ψ_3: The function $1/\sqrt{b \mp x}$, with $b = g^2 + 1$, has a series expansion in the form of Eq. (A18), but with the fraction $1/2$ inside the binomial coefficient replaced by $-1/2$. Therefore, by analogy with the result expressed in Eq. (A21), we have

$$\begin{aligned}\psi_3(t) \;=\; &\sum_{m=0}^{\infty} \binom{-\frac{1}{2}}{m} (\mp g)^m \left(g^2+1\right)^{\frac{1}{2}-m} \times \\ &\sum_{k=0}^{m} \binom{m}{k} \delta(2k\tau_c - t - m\tau_c),\end{aligned} \tag{A32}$$

which has the same convergence behavior as that of ψ_1.

- Ψ_4: The IFT of $\Psi_4 = 2\gamma x = 4\gamma g\cos(\omega\tau_c)$ is straightforward:

$$\psi_4 = 2\gamma g\delta(\tau_c - t) + 2\gamma g\delta(t + \tau_c). \tag{A33}$$

- Ψ_5: Explicitly, Eq. (A9) is

$$\Psi_5 = \frac{\sec^3(\omega\tau_c)}{(4g\gamma)^3}, \tag{A34}$$

where, following the same arguments as in the case of Ψ_2, the function $\sec^3(\omega\tau_c)$ can be expanded in a convergent series of the form of that in Eq. (A28), but with the fraction $-1/2$ inside the binomial coefficient replaced by $-3/2$. Therefore, by analogy with the result expressed in Eq. (A31), we have

$$\begin{aligned}\psi_5 \;=\; &\pm\frac{1}{(4g\gamma)^3}\sum_{m=0}^{\infty} \binom{-\frac{3}{2}}{m} (-1)^m \times \\ &\sum_{k=0}^{2m} \binom{2m}{k} (-1)^{k+m} 4^{-m} \delta(t + 2k\tau_c - 2m\tau_c).\end{aligned} \tag{A35}$$

- Ψ_6: Eq. (A10) can be written as

$$\Psi_6 = \frac{1}{1 - y(c)} \tag{A36}$$

where

$$y \equiv \frac{c}{x} = \frac{c}{2g\cos(\omega\tau_c)}, \tag{A37}$$

and c represents either c_1 or c_2, given by Eqs. (A11) and (A12), respectively. We wish to expand the function in Eq. (A36) into the power series

$$\sigma(c) \equiv \sum_{p=0}^{\infty} y^p(c) \tag{A38}$$

but this series converges only for

$$|y(c)| < 1. \tag{A39}$$

We now show that this convergence condition leads to a restriction on the allowable range of γ and g, but that this restriction does not limit the applicability of the IR to real listening configurations.

The inequalities in Eq. (A25), and the band division conditions in Eq. (56) and (58), imply that $x = 2g\cos(\omega\tau_c)$ is always positive in Branch-I bands and negative in Branch-II bands. Furthermore, we see from Eqs. (A11) and (A12) that, under the conditions in Eq. (A23), $c_1 \geq 0$ and $c_2 \leq 0$. Therefore, we have

$$y(c_1) = c_1/x \;\geq\; 0 \text{ in Branch-I bands,} \tag{A40}$$
$$y(c_1) = c_1/x \;\leq\; 0 \text{ in Branch-II bands,} \tag{A41}$$

and

$$y(c_2) = c_2/x \;\leq\; 0 \text{ in Branch-I bands,} \tag{A42}$$
$$y(c_2) = c_2/x \;\geq\; 0 \text{ in Branch-II bands.} \tag{A43}$$

If we define $\eta^+(c)$ and $\eta^-(c)$ to be the non-dimensional frequencies, $\omega\tau_c$, at which $y(c) = +1$ and $y(c) = -1$, respectively, we can, in light of the expressions above, restate the convergence condition in Eq. (A39) as:

$$\sigma(c_1) \text{ converges in Branch-I bands if } \phi \leq \eta^+(c_1) \tag{A44}$$

$$\sigma(c_1) \text{ converges in Branch-II bands if } \eta^-(c_1) \leq \pi - \phi \tag{A45}$$

and

$$\begin{aligned}&\sigma(c_2) \text{ converges in Branch-I bands if}\\ &\quad \phi_1 \le \eta^-(c_2)) \qquad \text{(A46)}\\ &\sigma(c_2) \text{ converges in Branch-II bands if}\\ &\quad \eta^+(c_2) \le \pi - \phi. \qquad \text{(A47)}\end{aligned}$$

Therefore, for $\sigma(c)$ to converge in Branch-I and Branch-II bands, all four inequalities must be satisfied. To express these convergence conditions explicitly (i.e., in terms of conditions on γ and g), we first set $y(c) = +1$ and $y(c) = -1$, and solve for $\eta^+(c)$ and $\eta^-(c)$, respectively, to find

$$\eta^+(c_1) = \cos^{-1}\left(\frac{f(g,\gamma)-1}{16g\gamma^2}\right) \tag{A48}$$

$$\eta^-(c_1) = \cos^{-1}\left(-\frac{f(g,\gamma)+1}{16g\gamma^2}\right) \tag{A49}$$

and

$$\eta^+(c_2) = \cos^{-1}\left(-\frac{f(g,\gamma)-1}{16g\gamma^2}\right) \tag{A50}$$

$$\eta^-(c_1) = \cos^{-1}\left(\frac{f(g,\gamma)+1}{16g\gamma^2}\right) \tag{A51}$$

where, for compactness, we have used the function $f(g,\gamma)$ defined as

$$f(g,\gamma) \equiv \sqrt{16\left(g^2+1\right)\gamma^2+1}.$$

Using these four explicit expressions, along with the definition of ϕ given by Eq. (60), we find that the inequalities in Eqs. (A44) and (A47) lead to the same explicit convergence condition:

$$\frac{f(g,\gamma)+7}{8(g^2+1)\gamma^2} \le 1; \tag{A52}$$

and the inequalities in Eqs. (A45) and (A46) lead to

$$\frac{f(g,\gamma)+9}{8(g^2+1)\gamma^2} \le 1. \tag{A53}$$

Since both of these inequalities need to be satisfied, and since the latter condition is more stringent than the former, we must satisfy the latter. We can finally state the condition for $\sigma(c)$ to converge in both Branch-I and Branch-II bands explicitly in terms of g and γ:

$$\frac{\sqrt{16\left(g^2+1\right)\gamma^2+1}+9}{8(g^2+1)\gamma^2} \le 1. \tag{A54}$$

This convergence condition is illustrated in the region plot of Fig. 9, where the black region denotes the values of g and γ for which the convergence condition is violated. It is clear that this restriction only slightly limits the range of allowable γ and g, and is not relevant to real listening geometries, where $g \simeq 1$.

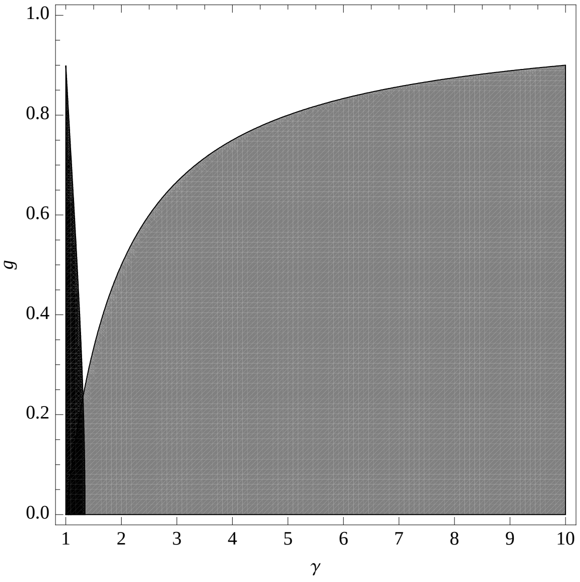

FIG. 9: Region plot showing the allowed values for g and γ (white). The black-shaded region is where the series convergence condition in Eq. (A54) is not satisfied, and the grey-shaded region is where the general condition in Eq. (52) is violated.

Aside from the series convergence condition above, γ must satisfy the general condition given by Eq. (52) (whose region of violation is shaded in grey in Fig. 9). Therefore, we combine both conditions in the following expression:

$$\max\left(\frac{\sqrt{5+\sqrt{5}}}{2\sqrt{g^2+1}}, 1\right) \le \gamma \le \frac{1}{1-g}, \tag{A55}$$

where the first argument of the max function comes from setting the left-hand side of the convergence condition in Eq. (A54) to 1, and solving for γ.

Now that we have found the convergence condition for the series in Eq. (A38), we can express Ψ_6 as that series and proceed to find its IFT. Replacing y and x in that series by their explicit values, we write

$$\Psi_6 = \sum_{p=0}^{\infty}\left(\frac{c}{2g}\right)^p \sec^p(\omega\tau_c). \tag{A56}$$

The $\sec^p(\omega\tau_c)$ term can be expanded in a convergent series of the same form as the series in Eq. (A28), but with the fraction $-1/2$ inside the binomial coefficient replaced by $-p/2$, and this leads to:

$$\Psi_6 = \sum_{p=0}^{\infty}\left(\frac{\pm c}{2g}\right)^p \sum_{m=0}^{\infty}\binom{-\frac{p}{2}}{m}(-1)^m \sin^{2m}(\omega\tau_c). \tag{A57}$$

Finally, recalling the finite sum in Eq. (A29), and the associated IFT in Eq. (A30), we arrive at the sought expression for the IFT of $\Psi_6(c)$:

$$\psi_6(c) = \sum_{p=0}^{\infty}\left(\frac{\pm c}{2g}\right)^p \sum_{m=0}^{\infty}\left(\begin{array}{c}-\frac{p}{2}\\ m\end{array}\right)(-1)^m \times \sum_{k=0}^{2m}\left(\begin{array}{c}2m\\ k\end{array}\right)(-1)^{k+m}4^{-m}\delta(t+2k\tau_c-2m\tau_c). \tag{A58}$$

The complete impulse response of the optimal XTC filter is assembled according to Eqs. (A13) to (A15), and is valid under the condition stated in Eq. (A55).

APPENDIX B: NUMERICAL VERIFICATION

The optimal XTC IR derived in the previous appendix was evaluated for the typical case of $g = .985$ and $\Gamma = 7$ dB, and plotted in Fig. 8. To verify the IR's validity and assess the effect of the number of terms in the series expansions, we calculated its Fourier transform and compared the resulting spectra to those obtained from the frequency-domain expressions of Section IV B. An example is shown in Fig. 10 for the Branch-I part of the XTC spectra (top panel) and envelope spectra (bottom panel).

We found that excellent agreement (within a few tenths of a dB) over all frequencies does not require taking more than the first few (5-10) terms of the infinite series in the expressions for all the ψ functions constituting the IR, except for ψ_1 and ψ_3, which, due to their slow convergence at and near the frequencies $\omega\tau_c = n2\pi$ ($n = 0, 1, 2, 3, 4, \dots$), require taking a larger number of terms. Approximating the infinite series in the expressions for ψ_1 and ψ_3 by a sum having a finite number of terms causes departures from the correct amplitude spectra at and near these frequencies. Due to the logarithmic frequency scale, the $n = 0$ departure appears as a slight bass roll-off in the first band (seen as the first dot in the first Branch-I band in the bottom panel of Fig. 10), and the $n \geq 1$ departures appear as narrow-band spikes (such as the one appearing as three vertical dots in the fifth band in the same plot). Increasing the number of terms in the series above 1000 reduces the amplitude of the bass roll-off and pushes it into the subwoofer frequency range, where XTC is not needed, and causes the $n \geq 1$ spikes to diminish in amplitude and frequency extent so as to become inaudible. (The XTC spectrum is more immune from the aforementioned departures, as seen in the top panel, because it is a ratio of left to right spectra.)

A similar analysis of the Branch-II part of the IR is not shown as the resulting spectra exhibit the same behavior as that described above.

[1] D. H. Cooper and J. L. Bauk, J. Audio Eng. Soc. **37**, 3 (1989).

[2] Throughout this paper, the words "recording" and "signal" are used interchangeably and are meant to also represent a live feed, or the HRTF-encoded signal for the artificial placement of sounds in a virtual acoustic space.

[3] M. A. Akeroyd *et al.*, J. Acoust. Soc. Am. **121**, 1056 (2007).

[4] D. B. Ward, J. Acoust. Soc. Am. **110**, 1195 (2001).

[5] A. Sæbø, Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2001.

[6] Throughout this paper, the word "level" is meant to represent, generally, a frequency-dependent amplitude.

[7] J. Blauert, *Spatial Hearing* (The MIT Press, Cambridge, MA, 2001).

[8] B. B. Bauer, J. Audio Eng. Soc. **9**, 148 (1961).

[9] B. S. Atal, M. Hill, and M. R. Schroeder, Apparent Sound Source Translator, US patent No. 3,236,949. Application filed Nov. 1962, granted Feb 1966.

[10] P. Damaske, J. Acoust. Soc. Am. **50**, 1109 (1970).

[11] An exception could be made for recordings in which the specific placement of sound images was made with full accounting for crosstalk during playback, e.g., the case of stereo soundfields constructed with pan-potted mono images and monitored over loudspeakers, common in popular music recording.

[12] C. Hugonnet and P. Walder, *Stereophonic Sound Recording* (John Wiley & Sons, Chichester, UK, 1998).

[13] While, as mentioned above, XTC levels above 20 dB are often required for robust front-back image disambiguation, the larger portion of the *direct* sound content in acoustic recordings, e.g., performed music, is of frontal origin and, with playback through frontal loudspeakers at modest XTC, is largely immune to such localization confusion.

[14] D. B. Ward and G. Elko, IEEE Signal Process. Lett. **6**, 106 (1999).

[15] J. J. Lopez and A. Gonzalez, IEEE Signal Process. Lett. **6**, 106 (1999).

[16] T. Takeuchi, P. A. Nelson, and H. Hamada, J. Acoust. Soc. Am. **109**, 958 (2001).

[17] J. Rose, P. Nelson, B. Rafaely, and T. Takeuchi, J. Acoust. Soc. Am. **112**, 1992 (2002).

[18] M. R. Bai and C. C. Lee, J. Acoust. Soc. Am. **120**, 1976 (2006).

[19] O. Kirekby, P. A. Nelson, and H. Hamada, J. Acoust. Soc. Am. **104**, 1973 (1998).

[20] O. Kirekby, P. A. Nelson, and H. Hamada, J. Audio Eng. Soc. **46**, 387 (1998).

[21] H. S. Kim, P. M. Kim, and H. B. Kim, ETRI J. **22**, 11 (2000).

[22] J. Yang, W. S. Gan, and S. E. Tan, Acous. Res. Lett. Online **4**, 47 (2003).

[23] M. R. Bai, C. C. Tung, and C. C. Lee, J. Acoust. Soc. Am. **117**, 2802 (2005).

[24] M. R. Bai and C. C. Lee, J. Sound & Vibration **117**,

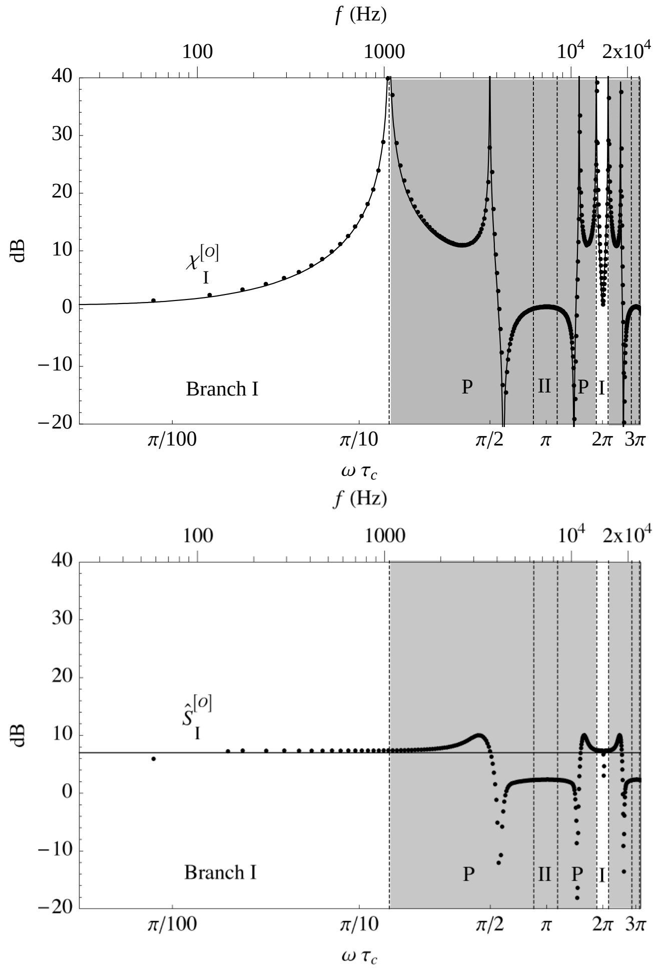

FIG. 10: XTC spectrum of optimal filter for Branch-I bands, $\chi_I^{[O]}(\omega)$, shown in top panel, and the associated envelope spectrum, $\hat{S}_I^{[O]}(\omega)$, shown in bottom panel. The small dots represent the spectra calculated by taking the Fourier transform of the Branch-I part of the IR derived in the previous appendix. (The IR is shown graphically in the top panel of Fig. 8.) Only the first 20 terms of the infinite series representing the ψ functions were taken, with the exception of the series for ψ_1 and ψ_3, for which the first 2500 terms were used. The hard curve in the top panel is the Branch-I XTC spectrum calculated directly from Eq. (64), and the horizontal line in the bottom panel is the Branch-I envelope spectrum $\hat{S}_I^{[O]}(\omega) = \gamma$, with $\Gamma = 7$ dB. (Other parameters are the same as for Fig. 2.) Since these spectra are valid only in Branch-I bands, all other bands are shaded in grey. (The vertical dashed lines represent the frequency bounds of the successive bands, and the branch numbers of the first five bands are given in the bottom half of each panel.)

2802 (2005).

[25] Y. Kim, O. Deille, and P. A. Nelson, J. Sound & Vibration **297**, 251 (2006).

[26] L. H. Kim, J. S. Lim, and K. M. Sung, IEICE Trans. Fundamentals **E85 A**, 2159 (2002).

[27] W. G. Gardner, *3-D Audio using Loudpseakers* (Kluwer Academic Publishers, Boston, 1998).

[28] Conversely, XTC control at frequencies where the interference at the ear is constructive requires attenuation instead of boosting (and implies a dynamic range gain, instead of loss). As was shown by [29, 30], and as will be reviewed in Section II C, these attenuations are not problematic as they correspond to frequencies where XTC control is most robust.

[29] T. Takeuchi and P. A. Nelson, J. Acoust. Soc. Am. **112**, 2786 (2002).

[30] P. A. Nelson and J. F. W. Rose, J. Acoust. Soc. Am. **118**, 193 (2005).

[31] B. Katz, *Mastering Audio* (Focal Press, Oxford, UK, 2002).

[32] T. Takeuchi and P. A. Nelson, J. Audio Eng. Soc. **5**, 981 (2007).

[33] A. Schuhmacher, J. Hald, K. B. Rasmussen, and P. C.Hansem, J. Acoust. Soc. Am. **113**, 114 (2003).

[34] H. Tokuno, O. Kirkeby, P. A. Nelson, and H. Hamada, IEICE Trans. Fundamentals **EE80-A**, 809 (1997).

[35] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).

[36] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, PA, 1998).

[37] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic Press, London, UK, 1993).

[38] M. Bellanger, *Digital Processing of Signals* (John Wiley & Sons, Chichester, UK, 2000).

[39] W. G. Gardner, J. Audio Eng. Soc. **43**, 127 (1995).

[40] A. Torger and A. Farina, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (PUBLISHER, New Paltz, New York, 2001), paper No. 198.

[41] M. R. Bai and C. C. Lee, EURASIP J. on Adv. in Sign. Process. **2007**, 1 (2006).

[42] This value for the effective inter-ear separation, Δr = 15 cm, is justified by the relatively small loudspeaker span, following the guidelines in Ref. [29], where it is reported that good correlation between the peak frequencies in the data calculated using a free-field model, and those measured with the KEMAR dummy head, can be obtained by taking an effective $\Delta r \simeq 13$ cm for low values of θ, and $\Delta r \simeq 25$ cm for large source azimuths. The larger value, which is much larger than the minimum distance between the entrances of the ear canals of the dummy head, reflects the effects of diffraction around the head.