Excerpts (with many pages omitted) from Chapter 5 of the Book *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*

Amazon link —>

Title: Binaural Audio Through Loudspeakers

Author: Choueiri, Edgar Y.

Publisher: Focal Press

Editors: Roginska, A., Geluso, P.

Date of Publication: October 12, 2017

Abstract: The most fundamental challenge in crosstalk-cancellation (XTC) filter design is dealing with the spectral coloration (tonal distortion) that XTC filters inherently impose on the sound emitted by the loudspeakers. The basic problem of XTC, the fundamental nature of the associated spectral coloration, its main features, its dependencies, and, ultimately, the formulation of a method for the practical design of optimal XTC filters that abate such tonal distortion, with minimal degradation of XTC performance, are the main subjects of this chapter.

Chapter 5

Binaural Audio Through Loudspeakers

Edgar Choueiri

Introduction

Background and Motivation

The ultimate goal of binaural audio with loudspeakers (BAL), also known as transauralization (Cooper & Bauck, 1989), is to reproduce, at each of the listener's eardrums, the sound pressure signals recorded on only the ipsilateral channel of a stereo signal. If the stereo signal[1] 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 channel of the stereo recording to the ipsilateral ear, and only to that ear, would ideally guarantee that the listener's ear-brain system receives the cues it needs to perceive an accurate three-dimensional reproduction of the recorded sound field. Since, with playback from two loudspeakers, each of the cues is also heard by the contralateral ear (crosstalk), accurate 3D audio reproduction through BAL requires an effective cancellation of this unintended crosstalk. Without such crosstalk cancellation (XTC), the ITD and ILD cues will inevitably be corrupted.

In addition to XTC, effective BAL requires an abatement of sound reflections in the listening room, since such reflections directly degrade the integrity of the binaural cues at the listener's ears (Damaske, 1971; Sæbø, 2001). While this problem can be somewhat alleviated through prescriptions that increase the ratio of direct to reflected sound, accurate sound localization through BAL has been shown to require XTC levels[2] above 20 dB (Parodi & Rubak, 2011b), which are difficult to achieve practically even under anechoic conditions (Akeroyd et al., 2007).

Therefore, it would seem that the goal stated in the first paragraph could be more naturally reached with binaural audio through headphones (or earphones), as both crosstalk and room reflections would be non-existent. However, with headphones or earphones, 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 three-dimensional image and to the perception that the sound (or some of its spectral components) is inside, or too close to, the listener's head (see Chapter 4 and Nicol, 2010, for a more thorough discussion of this issue).

Binaural playback through loudspeakers is largely immune to this head internalization of sound because, 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 (Atal, Hill & Schroeder, 1966; Bauer, 1961; Damaske, 1971 and Chapter 2).

Some applications of BAL, such as immersive virtual reality environments or scientific studies of spatial hearing, require binaural cues to be transmitted to a listener with a high degree of fidelity and reliability. Such transparency and robustness often require anechoic (or semi-anechoic) environments (or equivalently, high-directivity loudspeakers that abate the prominence of reflected sound), individualization of the XTC system for the listener and the playback set-up, precise matching of the listener's HRTF with that used in the recording, and either constraining the position of the listener's head in the area of equalization (the "sweet spot") (Akeroyd et al., 2007; Majdak, Masiero & Fels, 2013; Moore, Tew & Nicol, 2010; Parodi & Rubak, 2011b) or adding the complexity of a head-tracking system. However, in many less stringent applications, modest levels of XTC, even of a few dB over a limited range of frequencies, have the potential to significantly enhance the three-dimensional 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 corruption in the loudspeaker playback of recordings containing significant binaural cues.

This reduction of unintended corruption through XTC should also apply to the loudspeaker playback of most stereo recordings,[3] especially those made in real acoustic spaces, and even to recordings made using standard stereo microphone techniques without a dummy head, because these techniques all rely on preserving in the recording a good measure of the natural ITD and ILD cues needed for enhancing the accuracy of spatial localization and the realism of hall reverberation during playback (Hugonnet & Walder, 1997). We should therefore expect that effecting even a relatively low level of XTC in 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.[4]

Before addressing the most fundamental challenge in XTC filter design, we list some of the practical challenges encountered when implementing an effective XTC-enabled BAL playback system and refer to the literature that discusses effective solutions to these practical problems. As alluded to above, such XTC systems are typically sensitive to room reflections (Akeroyd et al., 2007; Damaske, 1971; Sæbø, 2001; Ward, 2001), require the use of specialized playback set-ups (Kirkeby, Nelson & Hamada, 1998a, b; Takeuchi & Nelson, 2002, 2007), and necessarily create a single restricted sweet spot in which the XTC is effective (Takeuchi, Nelson & Hamada, 2001; Ward & Elko, 1999; Xie, 2013, and references therein). Much research effort has been expended on how to relieve the latter constraint and has resulted in potential solutions, of varying degrees

of practicality, which include widening the sweet spot through the use of multiple loudspeakers (Bai, Tung & Lee, 2005; Takeuchi & Nelson, 2002; Yang, Gan & Tan, 2003) and/or elevated loudspeakers (Parodi & Rubak, 2010), providing XTC at multiple listening locations through the use of multiple loudspeaker pairs (Bauck & Cooper, 1996; Kim, Deille & Nelson, 2006), and dynamically moving the sweet spot to follow the location of the listener's head by tracking it with optical sensors (Gardner, 1998; Lentz, 2006; Mannerheim, 2008).

The most fundamental challenge in XTC filter design is dealing with the *tonal distortion* (spectral coloration)[5] that XTC filters inherently impose on the sound emitted by the loudspeakers. As we will show in the following sections, the level of tonal distortion depends on the location of the sound source in the sound field, and therefore cannot be corrected through equalization, especially for audio signals containing more than a single sound source. The basic problem of XTC, the fundamental nature of the associated tonal distortion, its main features, its dependencies, and, ultimately, the formulation of a method for the practical design of optimal XTC filters that abate such tonal distortion with minimal degradation of XTC performance are the main subjects of this chapter.

The Problem of XTC-Induced Tonal Distortion

Nature of the Problem

One main difficulty in implementing XTC is to reduce the artifice of crosstalk without adding an artifice of another kind: tonal distortion. 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 distances between an ear and 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 that ear of the listener might be constructive, destructive, or complementary (90° out of phase). At the frequencies for which the interference between in-phase recorded signals is destructive at the ears (or, alternatively, the frequencies for which the interference between out-of-phase signals is constructive), 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 (Takeuchi & Nelson, 2002).[6] As shown in the section "Benchmark: Perfect Crosstalk Cancellation," in the case of a *perfect* XTC filter (defined as one that theoretically yields, in a free-field or anechoic environment, an infinite XTC level over the entire audio band) for typical listening configurations, these level boosts can easily be in excess of 30 dB, and therefore amount to severe tonal distortion.

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 is the crosstalk cancelled, but the frequency spectrum is also reconstructed perfectly, i.e., with no tonal distortion.

As recognized by Takeuchi and Nelson (2002) and P. A. Nelson and Rose (2005), and as further discussed in the section "Benchmark: Perfect Crosstalk Cancellation," the frequencies at which the level boosts are required correspond to the frequencies at which the system inversion (the mathematical inversion of the system's transfer matrix, which leads to the XTC filter) is ill conditioned. As a result, XTC control becomes highly sensitive to errors at these frequencies, so that even a small error in the alignment of the listener's head in the real world would lead to a significant 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 level

boosts which must necessarily be imposed at these frequencies would be fully audible, even in the sweet spot, as coloration (tonal distortion).

Takeuchi and Nelson (2002) show that, even in an ideal world where the loudspeakers–listener alignment is perfect, this tonal distortion imposed at the loudspeakers would present three 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 dynamic range. Since even professional audio equipment is seldom designed to have more than a few dB headroom above the levels required to reproduce the full dynamic range of realistic sound pressures (Katz, 2002), in order to avoid clipping in the case of the "perfect" XTC filter defined above, the dynamic range of the program would need to be decreased by more than 30 dB (minus the headroom). This is particularly problematic, for instance, in the case of wide-dynamic-range audio recorded in 16 or 24 bits (see Chapter 2 and references to these early efforts in the Bibliography of this chapter).

Previous Work and Goals of This Chapter

The history of crosstalk cancellation extends back to the seminal work of Bauer, Atal, Hill and Schroeder in the early 1960s (see references to these early efforts in the Bibliography of this chapter) and has since progressed at a faster rate with the advent of digital audio for which XTC can be readily implemented through digital filtering. We shall not attempt to review this history here, nor the various methods of implementing XTC (which range from older techniques applied in the analog domain (Atal et al., 1966), to time-domain signal manipulation algorithms, such as the RACE algorithm (Glasgal, 2007), and FFT-based digital convolution with finite impulse response (FIR) filters (SreenivasaRao, Mahalakshmi & VenkataRao, 2012), and instead focus our discussion on the problem of tonal distortion in XTC filters.

Takeuchi and Nelson (2002) have developed a method that not only yields excellent measured XTC performance (see also Akeroyd et al., 2007; Takeuchi & Nelson, 2007), but also effectively solves the problem of tonal distortion. However, their method, called the "Optimal Source Distribution" (OSD), which is discussed in the section "Benchmark: Perfect Crosstalk Cancellation," requires the use of a minimum of four (but typically six) transducers positioned at various angles around the listener.

The problem of XTC-induced tonal distortion 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 chapter, we study this problem in the context of XTC optimization, which we define as the maximization of XTC performance for a desired tolerable level of tonal distortion or, equivalently, the minimization of tonal distortion for a desired XTC performance.

The ultimate goal of the discussion in this chapter is to describe the design of "optimal XTC filters" (called BACCH filters) that do not suffer from the following drawbacks inherent to regular XTC filters:

- D1: Severe tonal distortion to the sound heard by the listener, even if that listener is sitting in the intended sweet spot.
- D2: Useful XTC levels are reached only at limited frequency ranges of the audio band.
- D3: Severe dynamic range loss when the sound is processed through the XTC filter or processor (while avoiding distortion and/or clipping).

In particular, we use a free-field two-point-source model and address, analytically, the fundamental aspects of tonal distortion control through both constant-parameter (frequency-independent) and frequency-dependent regularization methods. The use of regularization in the design of XTC filters was proposed by Kirkeby and colleagues (1998) to make the inversion of the system transfer matrix better behaved, and has since seen widespread adoption in the field. Specifically, constant-parameter regularization has been employed to control ill-conditioning in the design of HRTF-based XTC filters (e.g., Akeroyd et al., 2007; Kirkeby et al., 1998; Majdak et al., 2013), and frequency-dependent methods have been employed to tame high- and low-frequency amplification due to measured-HRTF inversion (e.g., Kirkeby & Nelson, 1999; Moore et al., 2010) and to control the temporal extent of the XTC filters (e.g., Parodi & Rubak, 2010, 2011a). Regarding the issues of tonal distortion and dynamic range loss, Papadopoulos and Nelson (2010) used constant-parameter regularization to limit the dynamic range loss inflicted by XTC, and Bai et al. (2005) and Bai & Lee (2006a) employed frequency-dependent regularization to impose gain limits on the XTC filters.

In the section "Constant-Parameter Regularization," we show that while the technique of constant-parameter (non-frequency-dependent) regularization may alleviate some of drawback D3, it inherently introduces spectral artifice of its own (specifically, while reducing the amplitude of the spectral peaks in the inverted transfer matrix, constant-parameter regularization results in undesirable narrow-band artifacts at higher frequencies and a roll-off at lower frequencies at the loudspeakers) and does little to alleviate the other two drawbacks (D1 and D2).

A discussion of the fundamental aspects of frequency-dependent regularization in the section "Frequency-Dependent Regularization" will lead us to our ultimate goal: a method for designing "optimal XTC filters" called "BACCH filters." The method relies on calculating the frequency-dependent regularization parameter (FDRP) that results in a flat amplitude versus frequency response at the loudspeakers (as opposed to a flat amplitude versus frequency response at the ears of the listener, as in previous design methods), thus forcing XTC to be effected into the phase domain only and relieving the XTC filter from the drawbacks of audible tonal distortion and dynamic range loss. When the method is used with any effective optimization scheme, it results in XTC filters that yield optimal XTC levels over any desired portion of the audio band, impose no tonal distortion on the processed sound beyond the tonal distortion inherent in the playback hardware and/or loudspeakers, and causes no dynamic range loss. XTC filters designed with this method and used in the system are not only optimal but, due to their being free from drawbacks D1, D2, and D3, allow for a most natural and spectrally transparent 3D audio reproduction of binaural or stereo audio through loudspeakers.

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 are useful for evaluating and comparing the tonal distortion and performance of XTC filters, and conclude with the definition and discussion of a benchmark for such comparisons: the perfect XTC filter.

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 reflection from the head and pinnae of the listener or any other physical objects), and that the loudspeakers radiate like point sources.

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 Morse and Ingard (1986)

$$P(r, i\omega) = \frac{i\omega\rho_0 q}{4\pi}\frac{e^{-ikr}}{r},$$

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 $\rho_0 q/4\pi$, 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 Figure 5.1, the air pressure due to the two sources, under the above-stated assumptions, add up as

$$P_L(i\omega) = \frac{e^{-ikl_1}}{l_1}V_L(i\omega) + \frac{e^{-ikl_2}}{l_2}V_R(i\omega). \tag{5.1}$$

Similarly, at the right ear, we have

$$P_R(i\omega) = \frac{e^{-ikl_2}}{l_2}V_L(i\omega) + \frac{e^{-ikl_1}}{l_1}V_R(i\omega). \tag{5.2}$$

Here, l_1 and l_2 are the path lengths between either source and the ipsilateral and contralateral ears, respectively, as shown in that figure.

In order to maintain a connection with the relevant literature, we adopt the same nomenclature used by Kirkeby et al. (1998a, b), Takeuchi and Nelson (2002), and P. A. Nelson and Rose (2005). 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

$$\Delta l \equiv l_2 - l_1 \qquad \text{and} \qquad g \equiv l_1 / l_2 \tag{5.3}$$

as the path length difference and path length ratio, respectively. An inspection of the geometry illustrated in Figure 5.1 shows that $0 < g < 1$, and that the path lengths can be expressed as

$$l_1 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 - \Delta r l \, sin(\theta)}, \tag{5.4}$$

$$l_2 = \sqrt{l^2 + \left(\frac{\Delta r}{2}\right)^2 - \Delta r l \, sin(\theta)}, \tag{5.5}$$

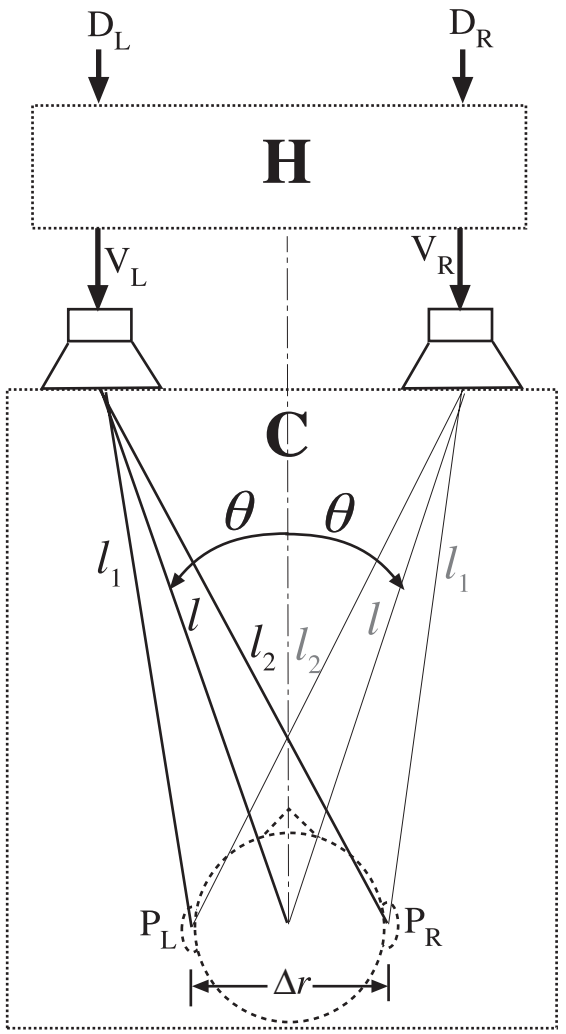

Figure 5.1 Geometry of the free-field two-point-source model. (All symbols are defined in the text.)

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 Figure 5.1, $2\theta = \Theta$ is the loudspeaker span. Note that for $l \gg \Delta r \sin(\theta)$, as in most loudspeaker-based listening set-ups, we have $g \simeq 1$. Another important parameter is the time delay,

$$\tau_c = \frac{\Delta l}{c_s}, \tag{5.6}$$

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

Using the above definitions, Equations (5.1) and (5.2) can be re-written in matrix form as

$$\begin{bmatrix} P_L(i\omega) \\ P_R(i\omega) \end{bmatrix} = \alpha \begin{bmatrix} 1 & ge^{-i\omega\tau_c} \\ ge^{-i\omega\tau_c} & 1 \end{bmatrix} \begin{bmatrix} V_L(i\omega) \\ V_R(i\omega) \end{bmatrix}, \tag{5.7}$$

where

$$\alpha = \frac{e^{-i\omega l_1/c_s}}{l_1}. \tag{5.8}$$

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 ensuring causality is discussed in the section "Metrics." The source vector $v = [V_L(i\omega), V_R(i\omega)]^T$ is obtained from the vector of "recorded" signals $d = [D_L(i\omega), D_R(i\omega)]^T$, through the transformation

$$\boldsymbol{v} = \boldsymbol{H}\boldsymbol{d}, \tag{5.9}$$

where

$$\boldsymbol{H} = \begin{bmatrix} H_{LL}(i\omega) & H_{LR}(i\omega) \\ H_{RL}(i\omega) & H_{RR}(i\omega) \end{bmatrix} \tag{5.10}$$

is the sought 2 × 2 filter matrix. Therefore, from Equation (5.7), we have

$$\boldsymbol{p} = \alpha \boldsymbol{C}\boldsymbol{H}\boldsymbol{d}, \tag{5.11}$$

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

$$\boldsymbol{C} \equiv \begin{bmatrix} 1 & ge^{-i\omega\tau_c} \\ ge^{-i\omega\tau_c} & 1 \end{bmatrix}, \tag{5.12}$$

which, like all matrices we deal with here, is symmetric due to the symmetry of the geometry.

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

$$\boldsymbol{p} = \alpha \boldsymbol{R}\boldsymbol{d}, \tag{5.13}$$

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

$$\boldsymbol{R} = \begin{bmatrix} R_{LL}(i\omega) & R_{LR}(i\omega) \\ R_{RL}(i\omega) & R_{RR}(i\omega) \end{bmatrix} \equiv \boldsymbol{C}\boldsymbol{H}. \tag{5.14}$$

Metrics

We now wish to define a set of metrics by which to judge the tonal distortion 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.

The responses of the system to a signal fed to only one (either left or right) of the two inputs, as heard at the ears, are called the "side images" of the system (i.e., either $\alpha\boldsymbol{R}\cdot[1,0]^T$ or $\alpha\boldsymbol{R}\cdot[0,1]^T$). We define our first coloration metric as the amplitude spectrum (to a factor α) of the side image at the ipsilateral ear, given by

$$E_{Si_{\|}}(\omega) \equiv \left| R_{LL}(i\omega) \right| = \left| R_{RR}(i\omega) \right|,$$

where the subscripts "si" and "||" stand for "side image" and "ipsilateral ear (with respect to the input signal)," respectively. Similarly, at the contralateral ear to the input signal (subscript "X"), we have the following side-image amplitude spectrum:

$$E_{Si_X}(\omega) \equiv \left| R_{RL}(i\omega) \right| = \left| R_{LR}(i\omega) \right|.$$

The response of the system to a signal split equally between left and right inputs, as heard at either ear, is called the "center image" of the system (i.e., $\alpha\boldsymbol{R}\cdot[1/2,1/2]^T$). We define another coloration metric as the amplitude spectrum of the center image, given by

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

where the subscript "ci" stands for "center image."

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

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

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

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

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 tonal distortion 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:

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

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

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

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 by Takeuchi and Nelson (2002).

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

$$\chi(\omega) \equiv \frac{|R_{LL}(i\omega)|}{|R_{RL}(i\omega)|} = \frac{|R_{RR}(i\omega)|}{|R_{LR}(i\omega)|} = \frac{E_{si_{\|}}(\omega)}{E_{si_X}(\omega)}.$$

The above definitions give us a total of eight metrics, ($E_{si\|}$, $E_{siˣ}$, E_{ci}, $S_{si\|}$, $S_{siˣ}$, S_{ci}, $\hat{S}$, and χ), all real functions of frequency, by which to evaluate and compare the tonal distortion and XTC performance of XTC filters.

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 the section "Background and Motivation," 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, Equation (5.13) requires that $\boldsymbol{R} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix, and thus, as per the definition of $\boldsymbol{R}$ in Equation (5.14), the P-XTC filter is simply the inverse of the system transfer matrix expressed in Equation (5.12), and can be expressed exactly:

$$\boldsymbol{H}^{[P]} = \boldsymbol{C}^{-1} = \frac{1}{1-g^2e^{-2i\omega\tau_c}}\begin{bmatrix} 1 & -ge^{-i\omega\tau_c} \\ -ge^{-i\omega\tau_c} & 1 \end{bmatrix}, \tag{5.15}$$

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

$$
\begin{aligned}
&E_{\mathrm{si}_{\parallel}}^{[P]} = 1; \quad E_{\mathrm{si}_X}^{[P]} = 0; \quad E_{\mathrm{ci}}^{[P]} = \frac{1}{2} \\
&S_{\mathrm{si}_{\parallel}}^{[P]}(\omega) = \left| \frac{1}{1 - g^2 e^{-2i\omega\tau_c}} \right| \\
&\qquad = \frac{1}{\sqrt{g^4 - 2g^2 \cos(2\omega\tau_c) + 1}}; \\
&S_{\mathrm{si}_X}^{[P]}(\omega) = \left| \frac{-g e^{-i\omega\tau_c}}{1 - g^2 e^{-2i\omega\tau_c}} \right| \\
&\qquad = \frac{g}{\sqrt{g^4 - 2g^2 \cos(2\omega\tau_c) + 1}}; \\
&S_{\mathrm{ci}}^{[P]}(\omega) = \frac{1}{2} \left| 1 - \frac{g}{g + e^{i\omega\tau_c}} \right| \\
&\qquad = \frac{1}{2\sqrt{g^2 + 2g \cos(\omega\tau_c) + 1}}; \\
&\hat{S}^{[P]}(\omega) = \max\left(\left| 1 - \frac{g}{g + e^{i\omega\tau_c}} \right|, \left| 1 + \frac{g}{e^{i\omega\tau_c} - g} \right| \right) \\
&\qquad = \max \left(\frac{1}{\sqrt{g^2 + 2g \cos(\omega\tau_c) + 1}}, \right. \\
&\qquad\qquad \left. \frac{1}{\sqrt{g^2 + 2g \cos(\omega\tau_c) + 1}} \right);
\end{aligned}
\tag{5.16}
$$

$$
\chi^{[P]}(\omega) = \infty. \tag{5.17}
$$

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 tonal distortion at the loudspeakers, we plot the $S^{[P]}(\omega)$ frequency responses expressed above in Figure 5.2 for a typical value of $g = 0.985$. Throughout this chapter, 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 $g = 0.985$ and $\tau_c = 68\ \mu$s (i.e., 3 samples at the "Red Book" CD sampling rate of 44.1 kHz), which would be the case, for instance, in a set-up with $\Delta r = 15$ cm, $l = 1.6$ m, and $\Theta = 18°$.

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.

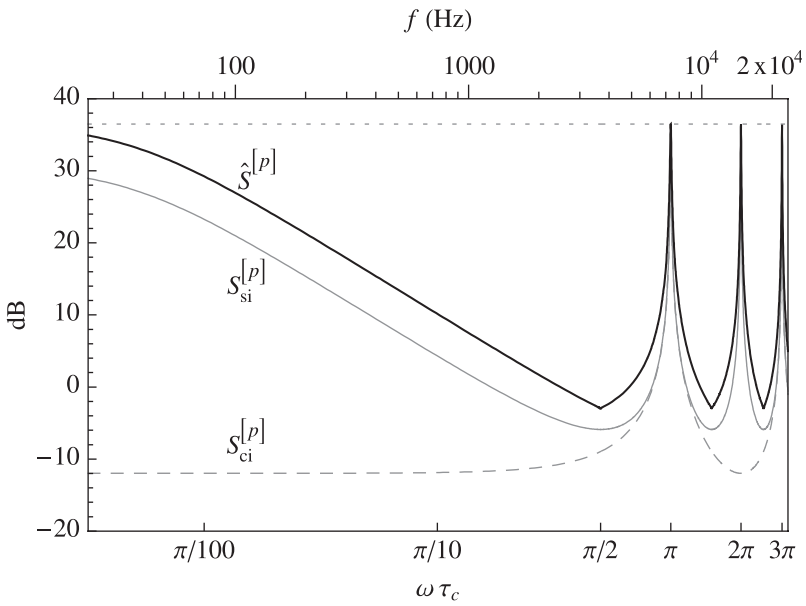

Figure 5.2 Perfect XTC filter frequency responses at the loudspeakers: amplitude envelope (heavy curve), side image (light solid curve), and central image (light dashed curve). The dotted horizontal line marks the envelope ceiling, which for this case (g = 0.985) is 36.5 dB. The non-dimensional frequency $\omega\tau_\chi$ 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 $S_{si_\parallel}^{[P]} \simeq S_{si^x}^{[P]}$ when $g \simeq 1$, these two spectra are shown as the single curve $S_{si^x}^{[P]}$.)

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:

$$
\begin{aligned}
S_{si_\parallel}^{[P]\uparrow} &= \frac{1}{1-g^2} \text{ at } \omega\tau_c = n\pi, \\
S_{si_\parallel}^{[P]\downarrow} &= \frac{1}{1-g^2} \text{ at } \omega\tau_c = (2n+1)\frac{\pi}{2}, \\
S_{si_x}^{[P]\uparrow} &= \frac{1}{1-g^2} \text{ at } \omega\tau_c = n\pi, \\
S_{si_x}^{[P]\downarrow} &= \frac{1}{1-g^2} \text{ at } \omega\tau_c = (2n+1)\frac{\pi}{2}, \\
S_{ci}^{[P]\uparrow} &= \frac{1}{2-2g} \text{ at } \omega\tau_c = (2n+1)\pi,
\end{aligned}
\tag{5.18}
$$

$$S_{\text{ci}}^{[P]\downarrow} = \frac{1}{2-2g} \text{at } \omega\tau_c = 2n\pi,$$

$$\hat{S}^{[P]\uparrow} = \frac{1}{1-g} \text{ at } \omega\tau_c = n\pi,$$

$$\hat{S}^{[P]\downarrow} = \frac{1}{\sqrt{1+g^2}} \text{ at } \omega\tau_c = (2n+1)\frac{\pi}{2}, \tag{5.19}$$

with n = 0,1,2,3,4,. . . .

For a typical listening set-up, $g \simeq 1$, say, our reference $g = 0.985$ case shown in Figure 5.2, the envelope peaks (i.e., $\hat{S}^{[P]\uparrow}$) correspond to a boost of

$$20 \log_{10}\left(\frac{1}{1-0.985}\right) = 36.5 \text{ dB}$$

(and the peaks in the other sp*ectra,* $S_{\text{sill}}^{[P]\uparrow} \simeq S_{\text{six}}^{[P]\uparrow} \simeq S_{\text{ci}}^{[P]\uparrow}$ 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 (Kirkeby et al., 1998b; Takeuchi & Nelson, 2002). 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 Equations (5.4)–(5.6), is achieved by increasing l and/or decreasing the loudspeaker span Θ, as is done in the so-called Stereo Dipole configuration described by Kirkeby, Nelson and Hamada (1998a, b), where $\Theta = 10°$), the low-frequency boost of the P-XTC filter would remain problematic.

As mentioned in the section "The Problem of XTC-Induced Tonal Distortion," the severe tonal distortion 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 ($E^{[P]}(\omega)$ = 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 C.

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 P. A. Nelson and Rose, 2005.) The condition number $\kappa(C)$ of the matrix C is given by

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

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

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

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

$$\begin{aligned}&\kappa^{\uparrow}(C) = \frac{1+g}{1-g} \text{ at } \omega\tau_c = n\pi,\\ &\kappa^{\downarrow}(C) = 1 \text{ at } \omega\tau_c = (2n+1)\frac{\pi}{2}\\ &\text{with } n = 0,1,2,3,4,\ldots,\end{aligned} \tag{5.20}$$

as was also reported by Ward and Elko (1999) and P. A. Nelson and Rose (2005) in terms of wavelengths, and by Takeuchi and Nelson (2002) in terms of the wave number. First, we note that the maxima 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 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., $\kappa^{\uparrow}(C) = 132.3$ for our typical case of $g = 0.985$) at the non-dimensional frequencies $\omega\tau_c = 0, \pi, 2\pi, 3\pi, \ldots$. As $g \rightarrow 1$, 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 tonal distortion 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 (2002, 2007), which effectively solves both the robustness and tonal distortion problem of the P-XTC filter by ensuring that the system operates always under conditions where $\kappa(C)$ is small. This can be done by allowing the loudspeaker span to be a function of the frequency. More specifically, after noting that typically $l \gg \Delta r$, 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

$$\omega\tau_c \simeq \frac{2\pi f\Delta\sin(\theta)}{c_s} \text{ for } l \gg \Delta r, \tag{5.21}$$

we can re-write the robustness condition (stated in Equation (5.20)) as

$$\Theta(f) \simeq 2\sin^{-1}\left(\frac{(2n+1)c_s}{4f\Delta r}\right),$$

with $n = 0,1,2,3,4,\ldots$

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 Takeuchi and Nelson (2002, 2007) and P. A. Nelson and Rose (2005), 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 tonal distortion that is heard by listeners inside and outside the sweet spot.

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. Hansen (1998) provides a detailed discussion of regularization methods in a general mathematical context, and others (e.g., Bai et al., 2005; Kirkeby & Nelson, 1999; Majdak et al., 2013; Parodi & Rubak, 2010) have demonstrated 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 tonal distortion or, equivalently, the minimization of tonal distortion for a desired minimum XTC performance.

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

$$\boldsymbol{H}^{[\beta]} = [\boldsymbol{C}^H \boldsymbol{C} + \beta \boldsymbol{I}]^{-1} \boldsymbol{C}^H, \tag{5.22}$$

where the superscript "H" denotes the conjugate transpose, 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. The pseudoinverse matrix $\boldsymbol{H}^{[\beta]}$ is the regularized filter, and the superscript "[β]" is used to denote constant-parameter regularization. The regularization stated in Equation (5.22) can be shown to correspond to a minimization of a cost function, $J(i\omega)$, where

$$J(i\omega) = e^H(i\omega)e(i\omega) + \beta v^H(i\omega)v(i\omega), \tag{5.23}$$

and the vector e represents a performance metric that is a measure of the departure from the signal reproduced by the perfect filter (Kirkeby et al., 1998; P. A. Nelson & Elliott, 1993). 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 $\beta > 0$, Equation (5.22) leads to an optimum, which corresponds to the least-squares 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 $S(\omega)$ spectra, at the price of a decrease in XTC performance at and near the frequencies where the system is ill conditioned.

Frequency Response

Using the explicit form for $\boldsymbol{C}$ given by Equation (5.12), in the last equation above, we find:

$$\boldsymbol{H}^{[\beta]} = \begin{bmatrix} H_{LL}^{[\beta]}(i\omega) & H_{LR}^{[\beta]}(i\omega) \\ H_{RL}^{[\beta]}(i\omega) & H_{RR}^{[\beta]}(i\omega) \end{bmatrix}, \tag{5.24}$$

where

$$\begin{aligned} H_{RR}^{[\beta]}(i\omega) &= H_{RR}^{[\beta]}(i\omega) \\ &= \frac{g^2 e^{4i\omega\tau_c} - (\beta+1)e^{2i\omega\tau_c}}{g^2 e^{4i\omega\tau_c} + g^2 - e^{4i\omega\tau_c}[(g^2+\beta)^2 + 2\beta + 1]}, \end{aligned} \tag{5.25}$$

$$\begin{aligned} H_{LR}^{[\beta]}(i\omega) &= H_{RL}^{[\beta]}(i\omega) \\ &= \frac{g^2 e^{i\omega\tau_c} - g(g^2+\beta)e^{3i\omega\tau_c}}{g^2 e^{4i\omega\tau_c} + g^2 - e^{2i\omega\tau_c}[(g^2+\beta)^2 + 2\beta + 1]}, \end{aligned} \tag{5.26}$$

The eight metric spectra we defined in the section "Metrics" become:

$$\begin{aligned} E_{\mathrm{si}_{\|}}^{[\beta]}(\omega) &= \frac{g^4 + \beta g^2 - 2g^2\cos(2\omega\tau_c) + \beta + 1}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\ E_{\mathrm{si}_X}^{[\beta]}(\omega) &= \frac{2g\beta\cdot\left|\cos(\omega\tau_c)\right|}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\ E_{\mathrm{ci}}^{[\beta]}(\omega) &= \frac{1}{2} - \frac{\beta}{2[g^2 + 2g\cos(2\omega\tau_c) + \beta + 1]}; \\ S_{\mathrm{si}_{\|}}^{[\beta]}(\omega) &= \frac{\sqrt{g^4 + 2(\beta+1)g^2\cos(2\omega\tau_c) + (\beta+1)^2}}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\ S_{\mathrm{si}_X}^{[\beta]}(\omega) &= \frac{g\sqrt{(g^2+\beta)^2 - 2(g^2+\beta)\cos(2\omega\tau_c) + 1}}{-2g^2\cos(2\omega\tau_c) + (g^2+\beta)^2 + 2\beta + 1}; \\ S_{\mathrm{ci}}^{[\beta]}(\omega) &= \frac{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}{2[g^2 + 2g\cos(2\omega\tau_c) + \beta + 1]}; \\ \hat{S}^{[\beta]}(\omega) &= \max\left(\frac{\sqrt{g^2 + 2g\cos(\omega\tau_c) + 1}}{g^2 + 2g\cos(2\omega\tau_c) + \beta + 1} \right. \end{aligned} \tag{5.27}$$

$$\chi^{[\beta]}(\omega) = \frac{g^4 + \beta g^2 - 2g^2\cos(2\omega\tau_c) + \beta + 1}{2g\beta \cdot \left|\cos(2\omega\tau_c)\right|}; \tag{5.28}$$

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 Figure 5.3 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).

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

$$\beta < \beta^* \equiv (g - 1)^2, \tag{5.29}$$

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

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

with n = 0,1,2,3,4, . . .

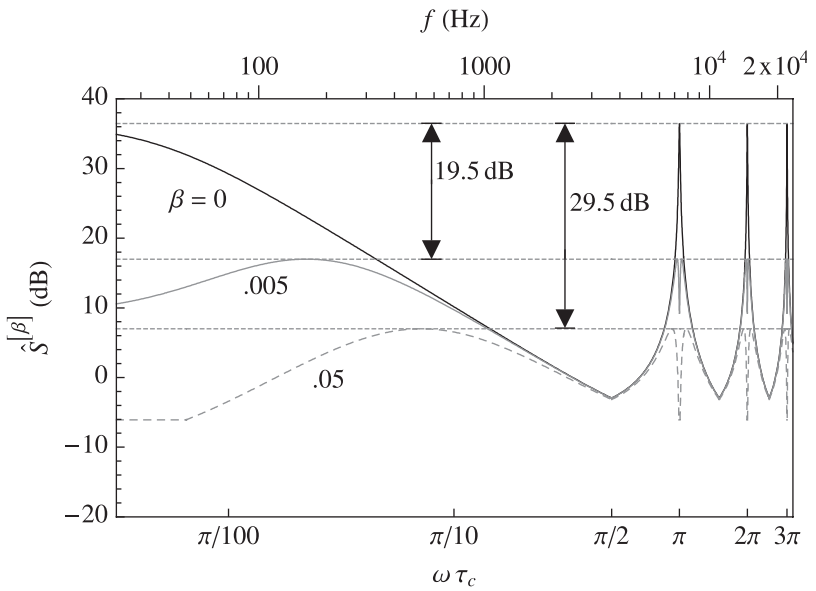

Figure 5.3 Effects of regularization on the envelope spectrum at the loudspeakers, $\hat{S}^{[\beta]}(\omega)$, showing peak attenuation and formation of doublet peaks as β is increased. (Other parameters are the same as for Figure 5.2.)

If the condition

$$\beta^* \leq \beta \ll 1 \tag{5.30}$$

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

$$\omega\tau_c = n\pi \pm \cos^{-1}\left(\frac{g^2 - \beta + 1}{2g}\right) \tag{5.31}$$

with $n = 0,1,2,3,4, \ldots,$

and have an amplitude

$$\hat{S}^{[\beta]\uparrow\uparrow} = \frac{1}{2\sqrt{\beta}}, \tag{5.32}$$

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 ($\beta = 0$) 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

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

For the typical case of $g = 0.985$ illustrated in Figure 5.3, we have $\beta^* = 2.25 \times 10^{-4}$, and for $\beta = 0.005$ 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 $g = 0.985$, we have $\beta^* = 2.25 \times 10^{-4}$ and $\Delta(\omega\tau_c) \simeq 0.225$ for $\beta = 0.05$.) 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 roll-off of typically many dB, as can be clearly seen in Figure 5.3. Therefore, constant-parameter 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 Figure 5.4.

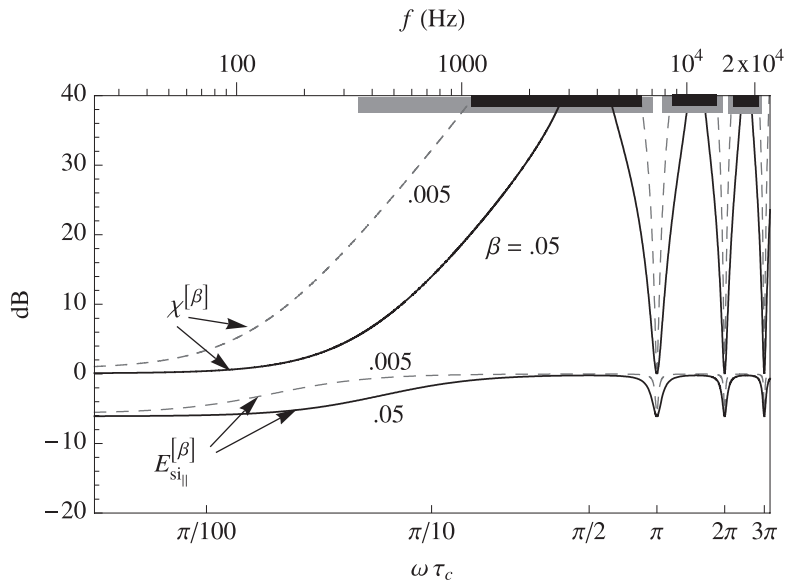

Figure 5.4 Effects of regularization on the crosstalk cancellation spectrum, $\chi^{[\beta]}(\omega)$ (top two curves), and the ipsilateral frequency response at the ear for a side image, $E^{[\beta]}_{si_{||}}(\omega)$ (bottom two curves). The black horizontal bars on the top axis mark the frequency ranges for which an XTC level of 20 dB or higher is reached with β = 0.05, and the grey bars represent the same for the case of β = 0.005. (Other parameters are the same as for Figure 5.2.)

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, . . .) and whose frequency extent widens with increasing regularization. For example, increasing β to 0.05 limits XTC of 20 dB or higher to the frequency ranges marked by black horizontal bars on the top axis of Figure 5.4, 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 tonal distortion peaks below a required level at the loudspeakers may be tolerated.

The $E^{[\beta]}_{si_{||}}(\omega)$ responses at the ears, shown as the bottom curves in Figure 5.4, depart only by a few dB from the corresponding P-XTC (i.e., β = 0) filter response (which is a flat curve at 0 dB). More precisely and generally, the maxima and minima of the $E^{[\beta]}_{si_{||}}(\omega)$ spectrum are given by:

$$E^{[\beta]\uparrow}_{si_{||}}(\omega) = \frac{g^2+1}{g^2+\beta+1} \text{ at } \omega\tau_c = (2n+1)\frac{\pi}{2};$$

$$E^{[\beta]\downarrow}_{si_{||}}(\omega) = \frac{g^4+(\beta-2)g^2+\beta+1}{g^4+2(\beta-1)g^2+(\beta+1)^2} \text{ at } \omega\tau_c = n\pi,$$

with n = 0,1,2,3,4, . . .

For the typical ($g = 0.985$) example shown in the figure, we have, for $\beta = 0.05$, and $E_{\mathrm{si}_{\|}}^{[\beta]\uparrow} = -0.2\mathrm{dB}$ $E_{\mathrm{si}_{\|}}^{[\beta]\uparrow} = -6.1\mathrm{dB}$, showing that even relatively aggressive regularization results in a tonal distortion at the ears that is quite modest compared to the tonal distortion 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 "bass boost") in the envelope spectrum at the loudspeakers, it typically results in undesirable narrow-band artifacts at higher frequencies and a roll-off 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 the section "Frequency-Dependent Regularization."

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

Impulse Response

We start by making the substitution $z = e^{2i\omega\tau c}$ in Equations (5.25) and (5.26) to get

$$H_{LL}^{[\beta]}(z) = H_{RR}^{[\beta]}(z) = \frac{z^2 g^2 - z(\beta+1)}{z^2 g^2 + g^2 - z\,[(g^2+\beta)^2 + 2\beta + 1]}, \tag{5.33}$$

$$H_{LR}^{[\beta]}(z) = H_{RL}^{[\beta]}(z) = \frac{z^2[gz^{-1/2} - g(g^2+\beta)z^{1/2}}{z^2 g^2 + g^2 - z^2[(g^2+\beta)^2 + 2\beta + 1]}. \tag{5.34}$$

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

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

where

$$a_1 = \frac{a - \sqrt{a^2 - 4g^4}}{2g^2}, \quad a_2 = \frac{a - \sqrt{a^2 - 4g^4}}{2g^2}, \tag{5.35}$$

and

$$a = \left(g^2 + \beta\right)^2 + 2\beta + 1. \tag{5.36}$$

We can then re-write Equations (5.33) and (5.34) as

$$H_{LL}^{[\beta]}(z) = H_{RR}^{[\beta]}(z) = \left[z - \frac{(\beta+1)}{g^2}\right] \times \left(\frac{1}{1 - a_1 z^{-1}}\right)\left(\frac{1}{z - a_2}\right), \tag{5.37}$$

$$H_{LR}^{[\beta]}(z) = H_{RL}^{[\beta]}(z) = \left[\frac{z^{-1/2} - (g^2+\beta)z^{1/2}}{g^2}\right] \times \left(\frac{1}{1 - a_1 z^{-1}}\right)\left(\frac{1}{z - a_2}\right). \tag{5.38}$$

Since $0 < g < 1$, and $\beta \geq 0$, we see from Equations (5.35) and (5.36) that $0 \leq a_1 < 1$ and $a_2 > 1$, and therefore $|a_1 z^{-1}| < 1$ and $a_2 > |z|$. This allows us to express the terms $1/(1 - a_1 z^{-1})$ 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

$$H_{LL}^{[\beta]}(z) = H_{RR}^{[\beta]}(z) = \left[z - \frac{(\beta+1)}{g^2} \right] \times \left(\sum_{m=0}^{\infty} a_1^m z^{-m} \right) \left(\sum_{m=0}^{\infty} - a_2^{-m-1} z^m \right), \tag{5.39}$$

$$H_{LR}^{[\beta]}(z) = H_{RL}^{[\beta]}(z) = \left[\frac{z^{-1/2} - (g^2 + \beta) z^{1/2}}{g^2} \right] \times \left(\sum_{m=0}^{\infty} a_1^m z^{-m} \right) \left(\sum_{m=0}^{\infty} - a_2^{-m-1} z^m \right). \tag{5.40}$$

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

$$\boldsymbol{h}^{[\beta]} = \begin{bmatrix} h_{LL}^{[\beta]}(t) & h_{LR}^{[\beta]}(t) \\ h_{RL}^{[\beta]}(t) & h_{RR}^{[\beta]}(t) \end{bmatrix}. \tag{5.41}$$

We do so by substituting back $e^{2i\omega\tau c}$ for z in Equations (5.39) and (5.40), and taking the inverse Fourier transform (IFT) to get

$$\begin{aligned} h_{LL}^{[\beta]}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{LL}^{[\beta]}(i\omega) e^{i\omega t} d\omega \\ &= h_{RR}^{[\beta]}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{RR}^{[\beta]}(i\omega) e^{i\omega t} d\omega \\ &\quad \left[\delta(1 + 2\tau_c) - \frac{\beta+1}{g^2} \delta(t) \right] * \psi(t), \end{aligned} \tag{5.42}$$

$$\begin{aligned} h_{LR}^{[\beta]}(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{LR}^{[\beta]}(i\omega) e^{i\omega t} d\omega \\ &= h_{LR}^{[\beta]}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H_{RL}^{[\beta]}(i\omega) e^{i\omega t} d\omega \\ &= \left[\frac{\delta(t - \tau_c)}{g} - \frac{g^2 + \beta}{g^2} \delta(t + \tau_c) \right] * \psi(t), \end{aligned} \tag{5.43}$$

where the asterisk ($*$) denotes the convolution operation, and $\psi(t)$ is the IFT of the product of the two series appearing in Equations (5.39) and (5.40), and is given by the following convolution of two trains of Dirac delta functions:

$$\psi(t) = \left(\sum_{m=0}^{\infty} a_1^m \delta(t - 2m\tau_c) \right) * \left(\sum_{m=0}^{\infty} -a_2^{-m-1} \delta(t + 2m\tau_c) \right). \tag{5.44}$$

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

The impulse response (IR) represented by Equations (5.42) and (5.43) is plotted in Figure 5.5 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 Equations (5.37) and (5.38) simplify to

$$H_{LL}^{[P]}(z) = H_{RR}^{[P]}(z) = \frac{1}{1 - a_1 z^{-1}}, \tag{5.45}$$

$$H_{LR}^{[P]}(z) = H_{RL}^{[P]}(z) = \frac{g z^{-1/2}}{1 - a_1 z^{-1}}, \tag{5.46}$$

from which, through the inverse Fourier transform, we recover the IR of the perfect XTC filter derived by Atal et al. (1966):

$$h_{LL}^{[P]}(t) = h_{RR}^{[P]}(t) = \sum_{m=0}^{\infty} a_1^m \delta(t - 2m\tau_c) \tag{5.47}$$

$$h_{LR}^{[P]}(t) = h_{RL}^{[P]}(t) = -g\delta(t - \tau_c) * \sum_{m=0}^{\infty} a_1^m \delta(t - 2m\tau_c) \tag{5.48}$$

where $a_1 = g^2$ (obtained by setting $\beta = 0$ in Equations (5.35) and (5.36)) 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 $a_1^m = (l_1 / l_2)^{2m}$ after a time $2m\tau_c$.

The physical significance of this impulse response has been discussed by P. Nelson et al. (1997) and Kirkeby et al. (1998b) who, along with Atal et al. (1966) before, recognized the recursive nature of XTC filters. Briefly, a physical appreciation of the perfect XTC 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 Equation (5.9), 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 Figure 5.5) 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, it then reaches the right ear with a slightly smaller amplitude but is cancelled there by a negative pressure pulse of equal amplitude (that was emitted a time l_1/c_s 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.

The effects of regularization on the XTC IR were recognized by Kirkeby et al. (1998), and can be gleaned from a comparison of the three panels of Figure 5.5. When β is finite, the IR has a "pre-echo" (non-causal) part, i.e., it extends in reverse time ($t < 0$), as shown in Figure 5.5. As can also be seen in that figure and inferred from Equation (5.44), 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 corresponds in the frequency domain to a spectrum with abated peaks.

Missing pages for copyright reasons.

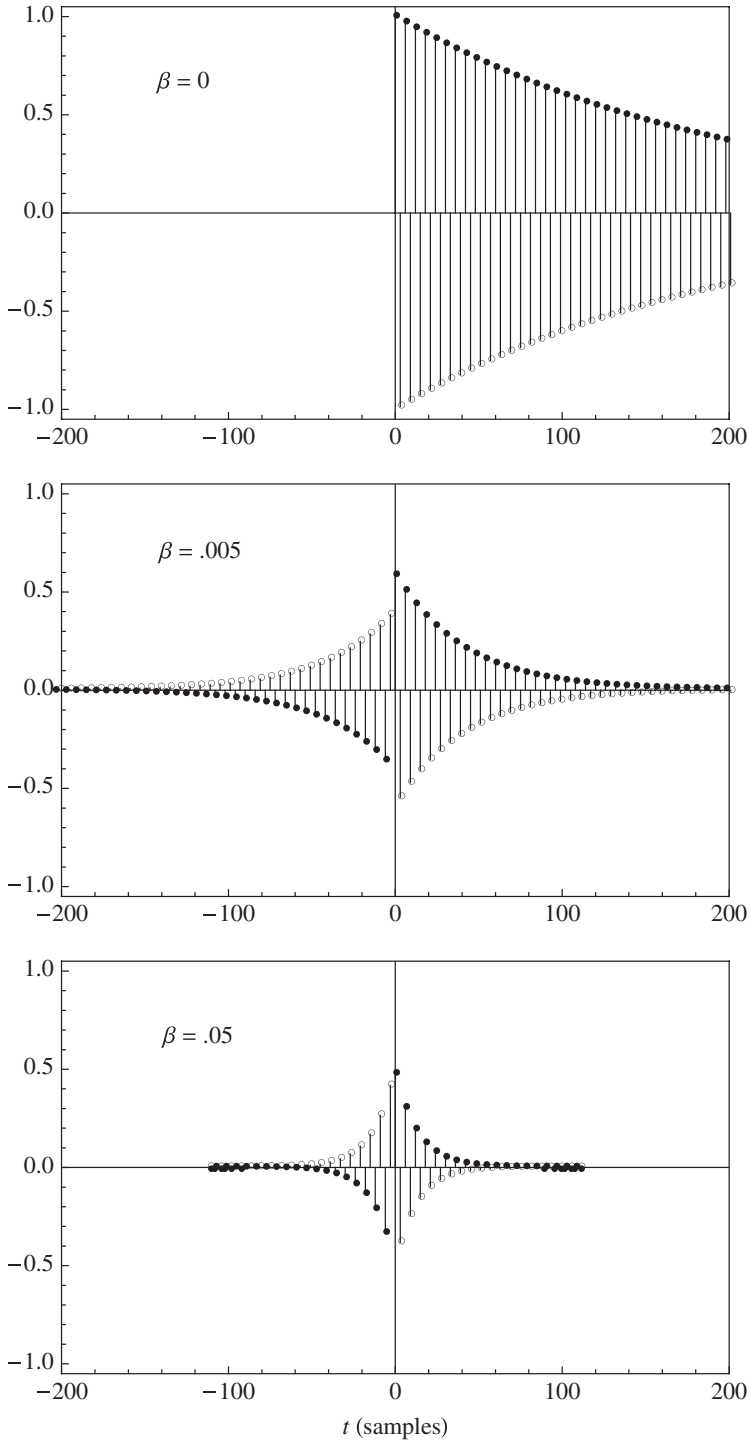

Figure 5.5 Impulse responses $h_{LL}^{[\beta]}(t) = h_{RR}^{[\beta]}(t)$ (filled circles) and $h_{LR}^{[\beta]}(t) = h_{RL}^{[\beta]}(t)$ $h_{RL}^{[\beta]}(t)$ (empty circles) for three values of β. (g = 0.985, τ_c = 3 samples)

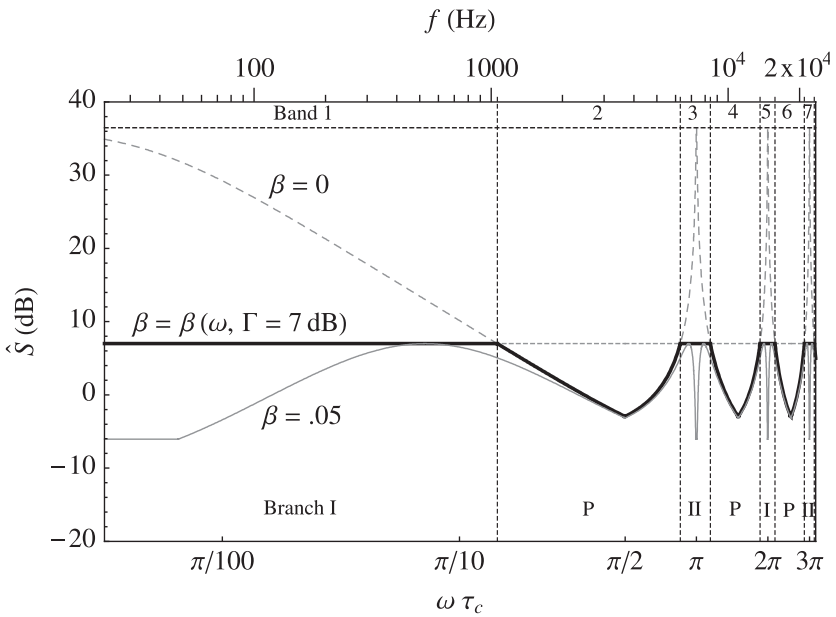

Figure 5.6 Envelope spectrum at the loudspeakers, $\dot{S}(\omega)$, for the case of frequency-dependent regularization with $\Gamma = 7$ dB (thick black curve) and for the corresponding reference case of $\beta = 0.05$ (grey curve). The benchmark case of the perfect XTC filter is also shown (dashed 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 Figure 5.2.)

Frequency Response

The amplitude envelope of the frequency response at the loudspeakers, given by Equation (5.49), was already shown in Figure 5.6. The other optimized metric spectra can be derived as follows:

$$Y_I^{[O]}(\omega) = Y^{[\beta_I(\omega)]}(\omega), \text{ for Branch–I bands;} \tag{5.60}$$

$$Y_I^{[O]}(\omega) = Y^{[\beta_{II}(\omega)]}(\omega), \text{ for Branch–II bands;} \tag{5.61}$$

$$Y_P^{[O]}(\omega) = Y^{[P]}(\omega), \text{ for Branch–P bands;} \tag{5.62}$$

where $Y(\omega)$ represents any of the eight metric spectra we defined in the section "Metrics," the superscript "$[O]$" denotes the sought optimized version of that metric spectrum, the subscripts "I," "II," and "P" denote each of the three branches, and the superscripts "$[\beta_I(\omega)]$" and "$[\beta_{II}(\omega)]$" denote regularization following the formulas for the regularized metric spectra in the section "Frequency Response," but with β taken to be frequency-dependent according to Equations (5.52) and (5.53).

For example, following the above hierarchical prescription, and using Equations (5.28), (5.52), (5.53), and (5.17), the optimized crosstalk cancellation spectrum becomes

$$\chi_{\mathrm{I,II}}^{[O]}(\omega) = \mp \frac{\gamma x(b \mp x) \mp b\sqrt{b \mp x}}{|x|(\gamma(b \mp x) - \sqrt{b \mp x}}, \tag{5.63}$$

$$\chi_{\mathrm{P}}^{[O]}(\omega) = \chi^{[P]}(\omega) = \infty, \tag{5.64}$$

where, for compactness, we have used the definitions $x \equiv 2g\cos(\omega\tau_c)$ and $b \equiv g^2 + 1$, 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, $E_{\mathrm{si}\|}(\omega)$, becomes

$$E_{\mathrm{si}_{\|\,\mathrm{I,II}}}^{[0]}(\omega) = \pm \frac{\gamma^2 x(b \mp x) \;\pm \gamma b\sqrt{b \mp x}}{(b \mp x) \;\pm\; 2\gamma x\sqrt{b \mp x}} \tag{5.65}$$

$$E_{\mathrm{si}_{\|\,\mathrm{P}}}^{[0]}(\omega) = E_{\mathrm{si}_{\|}}^{[P]}(\omega) = 1 \tag{5.66}$$

These spectra are plotted in Figure 5.7 where it is immediately clear from the $\chi(\omega)$ curves that frequency-dependent regularization yields a significant enhancement of XTC level over that

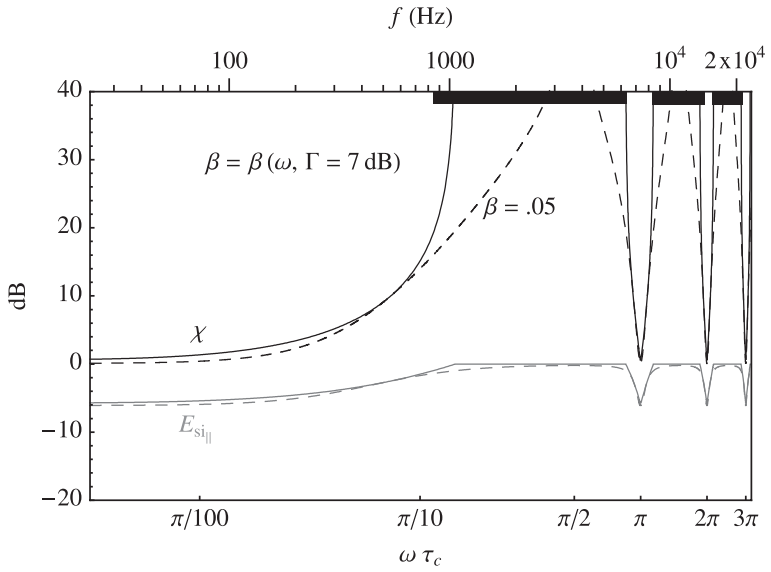

Figure 5.7 Crosstalk cancellation spectrum, $\chi(\omega)$ (black curves), and ipsilateral frequency response at the ear for a side image, $E_{\mathrm{si}\|}(\omega)$ (light curves), for the cases of frequency-dependent regularization (solid curves) and $\beta = 0.05$ (dashed curves). The frequency ranges 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 Figure 5.2.)

Missing pages for copyright reasons.

$$\psi_1 = \sum_{m-0}^{\infty}\binom{\frac{1}{2}}{m}(\mp g)^m\left(g^2+1\right)^{\frac{1}{2}-m} \times \sum_{k=0}^{m}\binom{m}{k}\delta\left(t-(2k-m)\tau_c\right),$$

$$\psi_2 = \pm\frac{1}{4g\gamma}\sum_{m-0}^{\infty}\binom{-\frac{1}{2}}{m}4^{-m} \times \sum_{k=0}^{2m}\binom{2m}{k}(-1)^k\,\delta\left(t+(2(\mathrm{m}-k)\tau_c\right),$$

$$\psi_3 = \sum_{m-0}^{\infty}\binom{-\frac{1}{2}}{m}(\mp g)^m\left(g^2+1\right)^{-\frac{1}{2}-m} \times \sum_{k=0}^{m}\binom{m}{k}\delta\left(t-(2k-m)\tau_c\right), \tag{5.75}$$

$$\psi_4 = 2g\gamma[\delta(t+\tau_c)+\delta(t+\tau_c)],$$

$$\psi_5 = \pm\frac{1}{(4g\gamma)^3}\sum_{m-0}^{\infty}\binom{-\frac{3}{2}}{m}4^{-m} \times \sum_{k=0}^{2m}\binom{2m}{k}(-1)^k\,\delta\left(t+(2(\mathrm{m}-k)\tau_c\right),$$

$$\psi_6(c) = \sum_{m-0}^{\infty}\left(\frac{\pm c}{2g}\right)^p\sum_{m=0}^{\infty}\binom{-\frac{p}{2}}{m}4^{-m} \times \sum_{k=0}^{2m}\binom{2m}{k}(-1)^k\,\delta\left(t+(2(\mathrm{m}-k)\tau_c\right),$$

with the constants c_1 and c_2 given by

$$c_1 = \frac{\sqrt{16\gamma^2(g^2+1)+1}\mp 1}{8\gamma^2}, \tag{5.76}$$

$$c_2 = \frac{-\sqrt{16\gamma^2(g^2+1)+1}\mp 1}{8\gamma^2}. \tag{5.77}$$

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

$$\max\left(\frac{\sqrt{5+\sqrt{5}}}{2\sqrt{g^2+1}},1\right) \le \gamma \le \frac{1}{1-g}, \tag{5.78}$$

which is shown graphically as a region plot in Figure 5.A.1 in Appendix A.

The impulse responses for Branch I and Branch II of this optimal filter are shown in Figure 5.8 for our typical case of $g = 0.985$ and $\tau_c = 3$ samples, and, along with the perfect filter IRs shown in the top panel of Figure 5.5, completely specify the optimal XTC filter.

Compared to the corresponding ($\beta = 0.05$) constant-parameter IRs in the bottom panel of Figure 5.5, the optimal XTC IRs shown in Figure 5.8 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 constant-parameter filters.

These IRs are difficult to interpret physically because 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 Figure 5.B.1, where the envelope spectrum

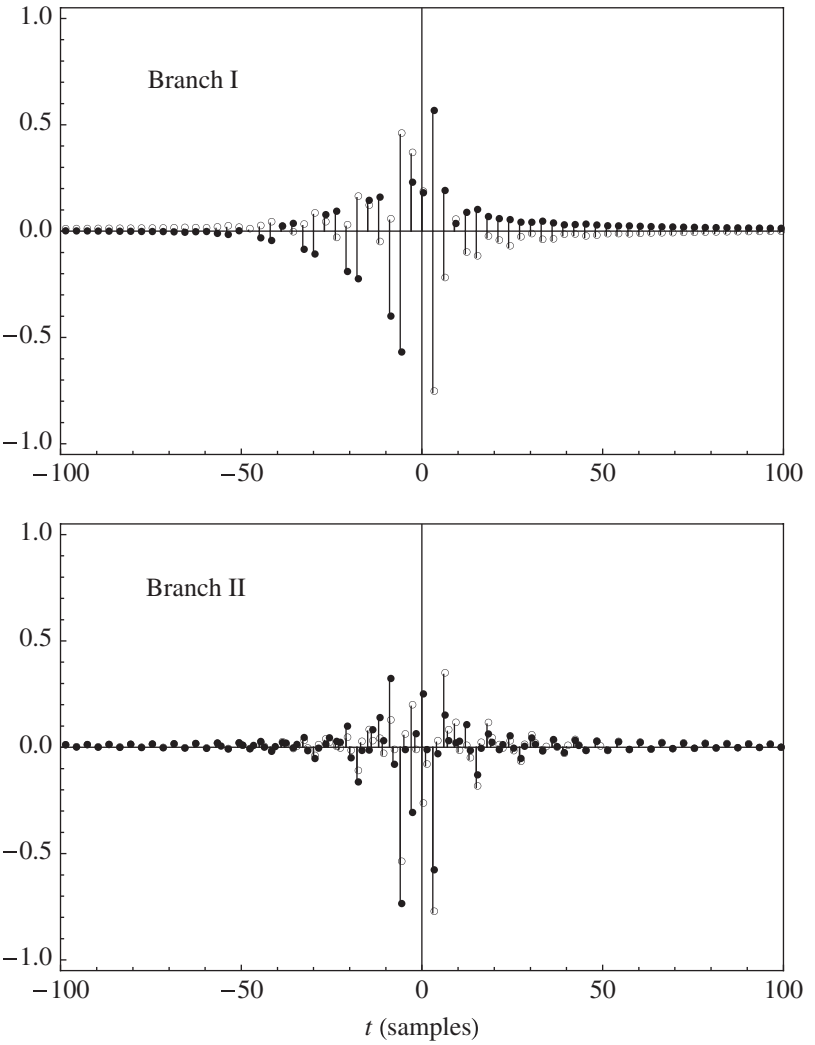

Figure 5.8 Impulse responses of the optimal XTC filter: $h_{LL}^{[O]}(t) = h_{RR}^{[O]}(t)$ (filled circles) and $h_{LR}^{[O]}(t) = h_{RL}^{[O]}(t)$ (empty circles), for Branch I (top panel) and Branch II (bottom panel). (Γ = 7 dB, g = 0.985, and τ_c = 3 samples, as in Figure 5.2.)

obtained from the Fourier transform of the Branch-I optimal IR is compared to the expected flat envelope spectrum, $\hat{S}_{\mathrm{I}}^{[O]}(\omega) = \gamma$. 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

Missing pages for copyright reasons.

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 Equations (5.42)–(5.44) (with $\beta = 0$), and the Branch-I and -II IRs from Equations (5.73)–(5.77). The loudspeakers source vector can then be calculated according to Equation (5.79), following the prescription given in the text preceding that equation, i.e., by appropriately convolving the 3-part IRs with the recorded stereo signal after having passed the latter through a multi-band crossover filter whose crossover frequencies are set to the band bounds given in Equations (5.55)–(5.58). The convolution operations can be carried out digitally, and in real-time if desired, using a digital convolution plugin. (Such software plugins often rely on FFT-based algorithms (e.g., Gardner, 1995) for fast convolution and have become readily available in the commercial and public domains for use as IR-based reverberation processors.)

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, the filter 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 2 × 2 IR matrix of the entire filter. Due to the symmetry of the filter, the other column of the IR matrix is obtained by simply flipping the two recorded outputs. This results in a single IR matrix, representing the entire three-branch multi-band filter, and simplifies any future application of Equation (5.79) to a simpler one (with no crossover filtering) in which the summation and indices are foregone.

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 as 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 upper 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 Equation (5.56), is $\pi - \varphi$), use Equation (5.21), and solve for θ, to get

$$\theta^* \simeq \sin^{-1}\left[\frac{c_s\left(\pi - \cos^{-1}\left[\frac{2\gamma^2 - 1}{2\gamma^2}\right]\right)}{2\pi f_c \Delta r}\right]. \tag{5.80}$$

A number of studies have suggested that XTC above a frequency of about 6 kHz is not critical or perhaps even necessary (Bai & Lee, 2007; Gardner, 1998; Majdak et al., 2013). Therefore, we set f_c equal 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. (Of course, to do so would require an additional 2-band crossover at f_c that precedes the one used to apply the XTC filter.)

It is relevant to mention in the context of loudspeaker span that keeping Θ small offers advantages that have been recognized since Kirkeby et al., (1998b) presented their analysis 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 (Bai & Lee, 2006b; Parodi & Rubak, 2010; Takeuchi et al., 2001). This effect 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 study by Bai and Lee (2006b) favored larger spans partly because increasing the span (while keeping the distance l fixed) lowers the value of g and consequently decreases the magnitude of the coloration peaks as well as the condition numbers. We do 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. The results of the study by Parodi and Rubak (2010), in which frequency-dependent regularization was employed subject to a 12 dB gain-limit on the XTC filters, seem to suggest that this is indeed the case.

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 chapter. 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 situation can be approached by decreasing the span angle as can be seen, for instance, in Figure 3.13 of Gardner (1998), 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 –2 dB) and flat (within 2 dB) for a small horizontal source azimuth ($\theta = 5°$), but increases and becomes less flat with increasing azimuths.

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 Equation (5.80), with $f_c \simeq 6$ kHz, and $\Delta r = 15$ cm,[7] we get $\theta = 9°$, which we take as half the loudspeaker span. From Equations (5.3)–(5.6), we then find $g = 0.985$ 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 chapter. The Branch-P and Branch-I IRs are therefore given by those shown in the top panels of Figure 5.5 and Figure 5.8, 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 Figure 5.6 and Figure 5.7, with the dimensional frequency read off the top axes of the plots, up to the cutoff frequency of 6 kHz. In

particular, we note that the XTC performance (top curve in Figure 5.7) exceeds 20 dB for a wide range of frequencies that extends from the 6 kHz cutoff down to 850 Hz, then drops off with decreasing frequency, reaching 5 dB at 290 Hz.

Individualized BACCH Filters

The BACCH Filter Design Method

Individualizing the BACCH filter to include the particular characteristics of the loudspeakers and the HRTF of the listener can lead to a significant enhancement of the realism of the 3D spatial imaging of binaural audio through loudspeakers.

We now describe the steps (shown schematically in Figure 5.9) of the technique (Choueiri, 2015) for designing such BACCH filters starting from the measured transfer function of a real listener in front of a pair of real loudspeakers.

- The starting point is a 2 × 2 impulse response measurement of the two loudspeakers using a binaural microphone in the ears of the listener. Such a measurement can be obtained through standard IR deconvolution using, for instance, the exponential sine-sweep technique (Farina, 2000, 2007). Each of the 4 impulse responses of this transfer function is FFTed to obtain the system's measured transfer matrix in the frequency domain (i.e., matrix C as in Equation (5.12)).
- In Step 1, the system's measured transfer matrix C is inverted numerically, using zero or a very small constant regularization parameter (large enough to avoid machine inversion problems) to obtain the corresponding perfect XTC filter, $\boldsymbol{H}^{[P]}$.
- In Step 2, the amplitude vs frequency response at the loudspeaker $\hat{S}^{[P]}$ is calculated and its lowest value (in dB) is taken to be Γ^*, then $\gamma^* = 10\Gamma^*/20$ is calculated.
- In Step 3, the frequency-dependent regularization parameter (FDRP), $\beta(\omega)$, that would result in a flat frequency response at the loudspeakers is calculated, so that $\hat{S}^{[\beta]}(\omega) = \text{constant} \leq \gamma^*$, thus forcing XTC to be caused by phase effects only.
- In Step 4, the FDRP thus obtained, $\beta(\omega)$, is used to calculate the pseudoinverse of the system's transfer matrix (e.g., according to Equation (5.22)), which yields the sought regularized optimal XTC filter $\boldsymbol{H}^{[\beta]}$ that has a flat frequency response at the loudspeakers. Finally, if needed for applying the resulting filter through a time-base convolution, as is often done in practical XTC implementation, a time domain version (impulse response) of the filter is obtained in the final step by simply taking the inverse Fourier transform of $\boldsymbol{H}[\beta]$.

It should be noted that in Step 3, if the FDRP is calculated so that $\hat{S}^{[\beta]}(\omega) = \text{constant} \leq \gamma^*$, the spectral flattening occurs for a side image (i.e., a sound panned to either channel and thus would be perceived by a listener to be located at or near the ipsilateral ear when the XTC level is sufficiently high). However, the same method can be used to flatten the response at the loudspeakers for an image that is not a pure side image by simply requiring that $S^{[\beta]}(\omega) = \text{constant} \leq \gamma^*$, where $S^{[\beta]}(\omega)$ is the XTC filter's frequency response for an image of source panned anywhere between the left and right channels. For instance, to flatten for a central image, we set $S_{ci}^{[\beta]}(\boldsymbol{\omega})$ (given, for instance, by the equation preceding Equation (5.27)) to a constant $\leq \gamma^*$, and proceed with the steps of the method as outlined above. In this context it is relevant to mention that for some

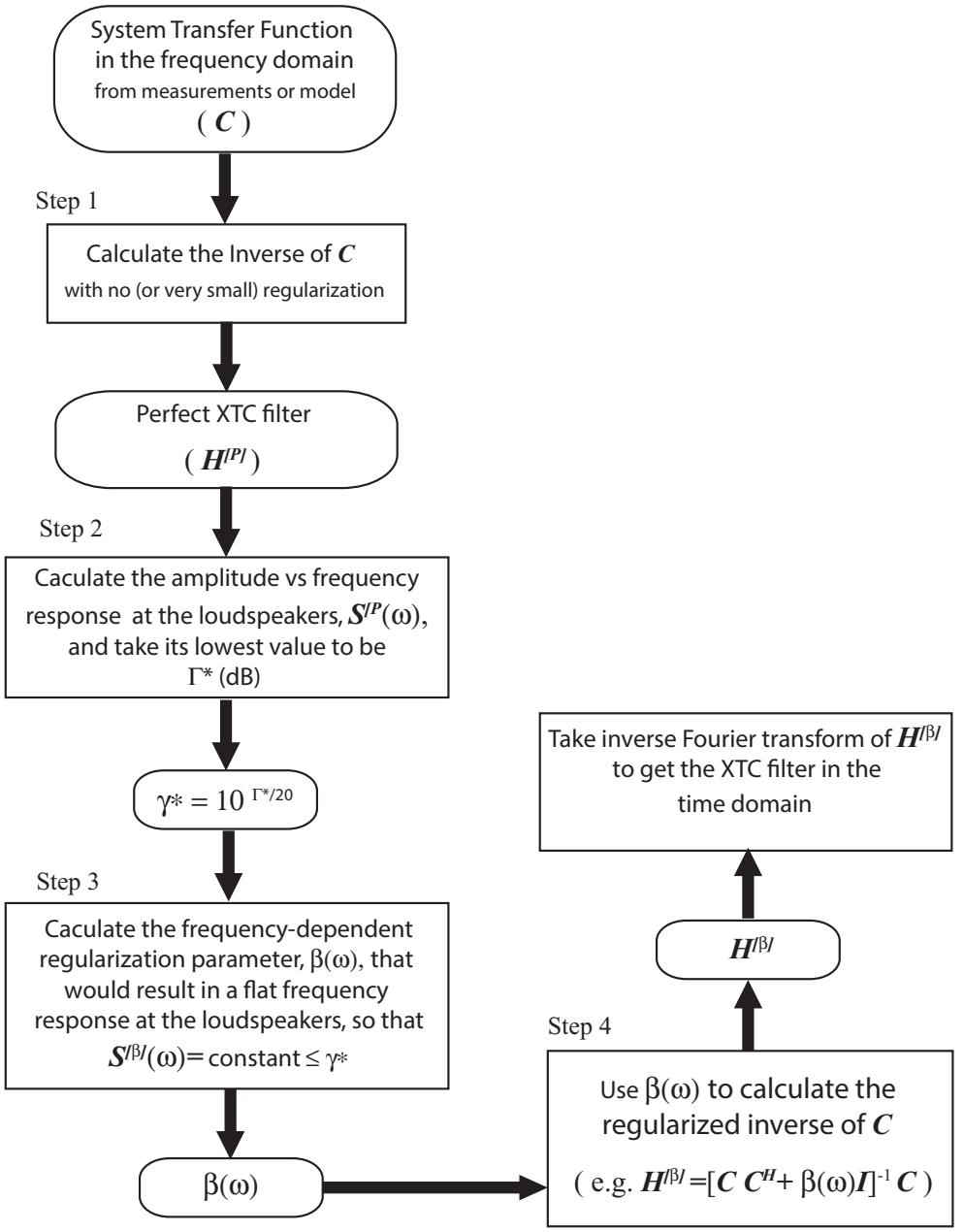

Figure 5.9 Flow chart illustrating the method for designing optimal XTC (i.e., BACCH) filters.

applications, for instance, pop music recording where the lead vocal audio is panned dead center, it might be desirable to flatten the response for a center image, i.e., $S_{ci}^{[\beta]}(\omega)$, (or an image of any other desired panning) in order to avoid coloration of that image. It should also be noted in that context that since $\hat{S}^{[\beta]}(\omega) \geq S^{[\beta]}(\omega)$, only flattening the side image (i.e., setting $S^{[\beta]}(\omega)$ = constant $\leq \gamma^*$) would result in no dynamic range loss. In other words, flattening for anything but the side image would incur a dynamic range loss that must be balanced by the benefit of a reduced tonal distortion for the desired panned image. For instance, for binaural recordings of real acoustic sound fields, which typically contain no dead-center panned images, flattening of the side image is advisable as it incurs no dynamic range loss.

Example Using a Measured Transfer Function

To illustrate the method described in the previous subsection, we give an example based on the transfer function of two loudspeakers in a room measured by microphones placed at the ear canal entrances of a dummy head (Neumann KU-100). The loudspeakers had a span of 60° at the listening position, which was about 2.5 m from each loudspeaker.

Figure 5.10 shows the four (windowed) measured impulse responses (IR) representing the transfer function in the time domain, and Figure 5.11 shows the spectra associated with the

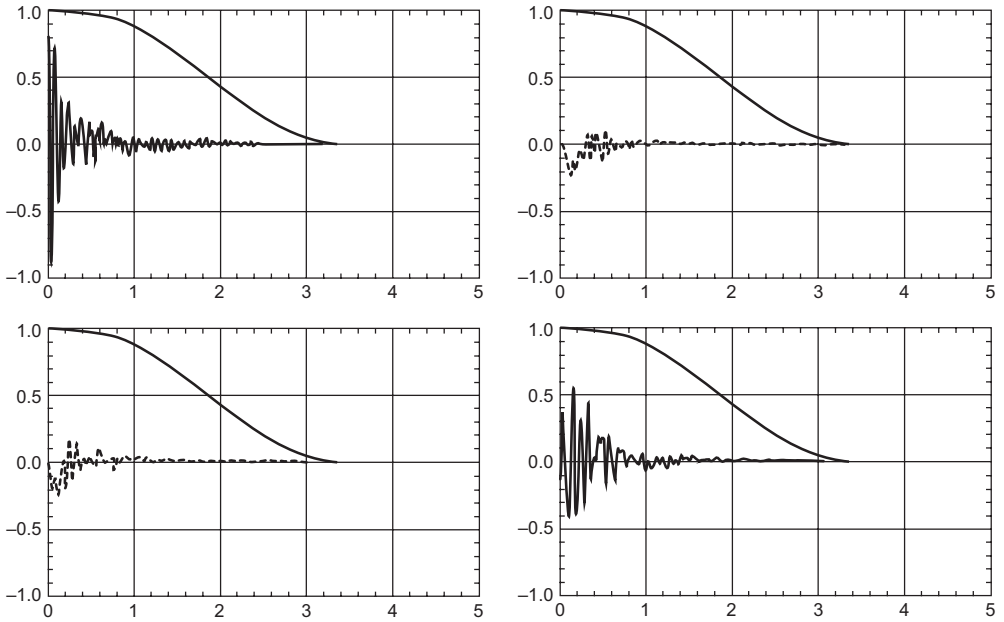

Figure 5.10 Four (windowed) measured impulse responses (IR) representing the transfer function in the time domain. The *x*-axis of each plot in that figure is time in ms, and the *y*-axis is the normalized amplitude of the measured signal. The top left plot shows the IR of the left loudspeaker measured at the left ear of the dummy head, and the bottom left plot shows the IR of the left loudspeaker measured at the right ear of the dummy head. The top right plot is the IR of the right speaker–left ear transfer function and the bottom plot is the IR of the right speaker–right ear transfer function.

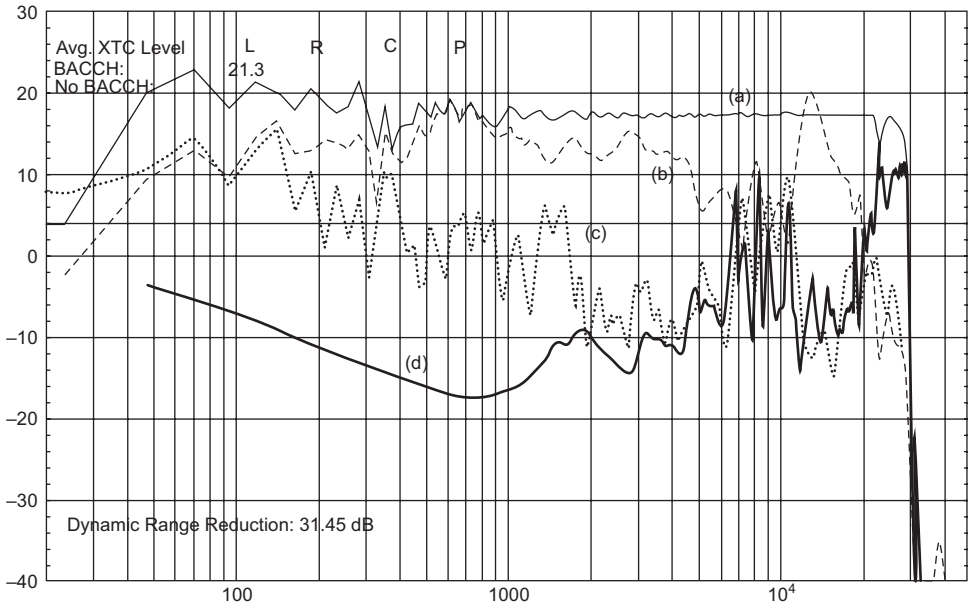

Figure 5.11 Measured spectra associated with the perfect XTC filter for the measured transfer function shown in Figure 5.10. The four curves represent: (a) the frequency response at the left (ipsilateral) ear; (b) the frequency response C_{LL} that corresponds to the left speaker–left ear transfer function; and (c) the frequency response measured at the right (contralateral) ear, E_{si_x}.

perfect XTC filter. The (b) curve in Figure 5.11 is the frequency response C_{LL} that corresponds to the left speaker-left ear transfer function in the frequency domain obtained by panning the test sound completely to the left channel. The ripples in that curve above 5 kHz are due to the HRTF of the head and the left ear pinna. The other curves in Figure 5.11 are the measured frequency responses associated with the perfect XTC filter, that is, an XTC filter obtained by inverting the transfer function with essentially no regularization ($\beta = 10^{-5}$). In particular, the (d) curve is the response at the left loudspeaker, $\hat{S}^{[\beta]}(\omega)$, and shows a dynamic range loss of 31.45 dB (difference between the maximum and minimum in that curve). The (a) curve is the frequency response at the left (ipsilateral) ear, E_{si}, which, as expected from a perfect XTC filter, is essentially flat over the entire audio band. The faint grey curve labeled (c) is the corresponding frequency response measured at the right (contralateral) ear, E_{six}, and shows significant attenuation with respect to the (c) curve due to XTC. The difference in amplitude between the (a) curve and (c) red curve, linearly averaged over frequencies, is the average XTC level, which for this case is 21.3 dB.

We contrast these curves with those curves in Figure 5.12, which shows the responses due to a filter designed in accordance with the BACCH filter design method.

By design, the curve labeled (d) in that plot, representing $\hat{S}^{[\beta]}(\omega) \geq S^{[\beta]}(\omega)$, the response at the left loudspeaker, is completely flat over the entire audio spectrum. Consequently, the frequency

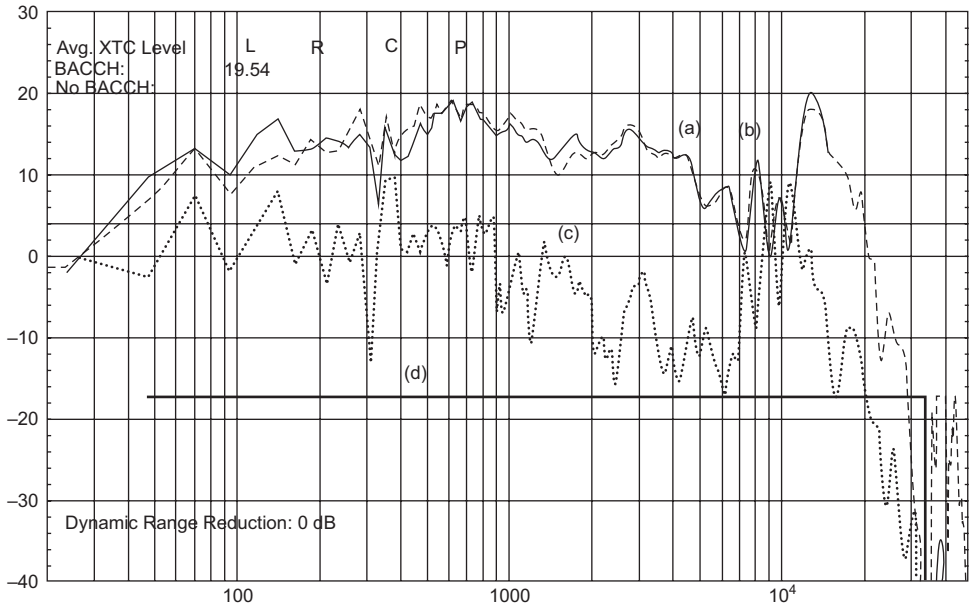

Figure 5.12 Measured spectra associated with the optimal XTC (BACCH) filter for the measured transfer function shown in Figure 5.10. The curves represent the frequency responses defined in the caption of Figure 5.11.

response at the left ear, curve (a), matches very well the corresponding measured system transfer function, C_{LL}, curve (b). Since $\hat{S}^{[\beta]}(\omega) \geq S^{[\beta]}(\omega)$ is flat, there is no dynamic range loss associated with this filter. The average XTC level for this filter (obtained by taking the linear average of the difference between the (a) and (c) curves) is 19.54 dB, which is only 1.76 dB lower than the XTC level obtained with the perfect filter, testifying to the optimal nature of the regularized filter.

In sum, the filter designed with the method described above imposes no audible coloration to the sound of the playback system, has no dynamic range loss, and yields an XTC level that is essentially the same as that of a perfect XTC filter.

Conclusions

Three-dimensional 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 tonal distortion 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 free-field two-point-source model, we showed that constant-parameter regularization, which has been used previously to design HRTF-based XTC systems, can lower these peaks, but also produces a bass roll-off and high-frequency artifacts in the filter's frequency response. Furthermore, we demonstrated that constant-parameter 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 the audio spectrum to be divided into a hierarchical set of adjacent frequency bands, each of which belongs 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.

The analytical XTC filters we 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 reflective 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.

We concluded with a discussion of a method for designing optimal individualized (HRTF-based) XTC filters (BACCH filters) that impose no audible coloration to the sound of the playback system, have no dynamic range loss, and yield the high XTC level attainable from a perfect XTC filter.

Notes

1 Throughout this chapter, 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.

2 Throughout this chapter, the word "level" is meant to represent, generally, a frequency-dependent amplitude.

3 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 sound fields constructed with pan-potted mono images and monitored over loudspeakers, common in popular music recording.

4 While it has been shown that reliable discrimination of frontal and rear images requires highly controlled playback and individualized XTC systems (Majdak et al., 2013), 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 levels of XTC, is largely immune to such localization confusion.

5 We use the terms "spectral coloration" and "tonal distortion" interchangeably.

6 On the other hand, at and near the frequencies for which the interference between in-phase (or out-of-phase) signals is complementary at the ears, XTC control requires slight attenuation instead of boosting (and implies a dynamic range gain, instead of loss). As shown by Takeuchi and Nelson (2002) and P. A. Nelson and Rose (2005), and as is reviewed in the section "Benchmark: Perfect Crosstalk Cancellation," these attenuations are not problematic as they correspond to frequencies where XTC control is most robust.

7 This value for the effective inter-ear separation, Δr = 15 cm, is justified by the relatively small loudspeaker span, following the guidelines of Takeuchi and Nelson (2002), who 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.

Bibliography

Akeroyd, M. A., Chambers, J., Bullock, D., Palmer, A. R., Summerfield, A. Q., Nelson, P. A., & Gatehouse, S. (2007). The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. *The Journal of the Acoustical Society of America*, *121*(2), 1056–1069. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/121/2/10.1121/1.2404625 doi: http://dx.doi.org/ 10.1121/1.2404625

Atal, B., Hill, M., & Schroeder, M. (1966, February 22). *Apparent Sound Source Translator*. Retrieved from www.google.com/patents/ US3236949. US Patent 3,236,949.

Bai, M. R., & Lee, C.-C. (2006a). Development and implementation of cross-talk cancellation system in spatial audio reproduction based on subband filtering. *Journal of Sound and Vibration*, *290*(3–5), 1269–1289. Retrieved from www.sciencedirect.com/science/ article/pii/S0022460X05003421 doi: http://dx.doi.org/10.1016/j.jsv.2005.05.016

Bai, M. R., & Lee, C.-C. (2006b). Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction. *The Journal of the Acoustical Society of America*, *120*(4), 1976–1989. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/120/4/10.1121/1.2257986 doi: http://dx.doi.org/ 10.1121/1.2257986

Bai, M. R., & Lee, C.-C. (2007). Subband approach to bandlimited crosstalk cancellation system in spatial sound reproduction. *EURASIP Journal of Advanced Signal Processing*, *2007*(071948), 1–9.69.

Bai, M. R., Tung, C.-W., & Lee, C.-C. (2005). Optimal design of loudspeaker arrays for robust cross-talk cancellation using the taguchi method and the genetic algorithm. *The Journal of the Acoustical Society of America*, *117*(5), 2802-2813. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/117/5/10.1121/1.1880852 doi: http://dx.doi.org/ 10.1121/1.1880852

Bauck, J., & Cooper, D. H. (1996). Generalized transaural stereo and applications. *Journal of Audio Engineering Society*, *44*(9), 683–705. Retrieved from http:// www.aes.org/e-lib/browse.cfm?elib=7888

Bauer, B. B. (1961). Stereophonic earphones and binaural loudspeakers. *Journal of Audio Engineering Society*, *9*(2), 148–151. Retrieved from www.aes.org/e-lib/ browse.cfm?elib=471

Bellanger, M. (2000). *Digital Processing of Signals: Theory and Practice*. Chichester, UK: John Wiley & Sons.

Choueiri, E. (2015). *Spectrally Uncolored Optimal Crosstalk Cancellation for Audio Through Loudspeakers*. Retrieved from www.google.com/ patents/WO2012036912A1?cl=en. International Patent Application No. PCT/US2011/050181, Granted November 18, 2015 under Patent No. 2612437.

Cooper, D. H., & Bauck, J. L. (1989). Prospects for transaural recording. *Journal of Audio Engineering Society*, *37*(1-2), 3–19. Retrieved from www.aes.org/e-lib/ browse.cfm?elib=6108

Damaske, P. (1971). Head-related two-channel stereophony with loudspeaker reproduction. *The Journal of the Acoustical Society of America*, *50*(4B), 1109-1115. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/50/4B/10.1121/1.1912742 doi: http://dx.doi.org/ 10.1121/1.1912742

Farina, A. (2000). Simultaneous measurement of impulse response and distorsion with a swept-sine technique. *Proceedings of the 108th Audio Engineering Society Convention*. Paris.

Farina, A. (2007). Advancements in impulse response measurements by sine sweeps. *Proceedings of the 122nd Audio Engineering Society Convention*. Vienna.

Gardner, W. G. (1995). Efficient convolution without input-output delay. *Journal of Audio Engineering Society*, *43*(3), 127–136. Retrieved from www.aes.org/ e-lib/browse.cfm?elib=7957

Gardner, W. G. (1998). *3-D Audio Using Loudspeakers*. Boston, MA: Kluwer Academic Publishers.

Glasgal, R. (2007). 360 degrees localization via 4. x RACE processing. *Proceedings of the 12rd Audio Engineering Society Convention*. Vienna.

Hansen, P. C. (1998). *Rank-deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Hugonnet, C., & Walder, P. (1997). *Stereophonic Sound Recording: Theory and Practice*. Chichester, UK: John Wiley & Sons.

Katz, B. (2002). *Mastering Audio: The Art and the Science* (pp. 61–74). Oxford, UK: Focal Press.

Kim, Y., Deille, O., & Nelson, P. (2006). Crosstalk cancellation in virtual acoustic imaging systems for multiple listeners. *Journal of Sound and Vibration*, *297*(1–2), 251–266. Retrieved from www.sciencedirect.com/science/article/pii/S0022460X06002884 doi: http:// dx.doi.org/10.1016/j.jsv.2006.03.042

Kirkeby, O., & Nelson, P. A. (1999). Digital filter design for inversion problems in sound reproduction. *Journal of Audio Engineering Society*, *47*(7-8), 583–595. Retrieved from www.aes.org/e-lib/browse.cfm?elib=12098

Kirkeby, O., Nelson, P. A., & Hamada, H. (1998a). Local sound field reproduction using two closely spaced loudspeakers. *The Journal of the Acoustical Society of America*, *104*(4), 1973–1981. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/ 104/4/10.1121/1.423763 doi: http://dx.doi.org/10.1121/1.423763

Kirkeby, O., Nelson, P. A., & Hamada, H. (1998b). The "stereo dipole": A virtual source imaging system using two closely spaced loudspeakers. *Journal of Audio Engineering Society*, *46*(5), 387–395. Retrieved from www.aes.org/ e-lib/browse.cfm?elib=12148

Kirkeby, O., Nelson, P. A., Hamada, H., & Orduna-Bustamante, F. (1998, March). Fast deconvolution of multichannel systems using regularization. *Speech and Audio Processing, IEEE Transactions On*, *6*(2), 189–194. doi: 10.1109/89.661479

Lentz, T. (2006). Dynamic crosstalk cancellation for binaural synthesis in virtual reality environments. *Journal of Audio Engineering Society*, *54*(4), 283–294. Retrieved from www.aes.org/e-lib/browse.cfm?elib=13677

Majdak, P., Masiero, B., & Fels, J. (2013). Sound localization in individualized and non-individualized crosstalk cancellation systems. *The Journal of the Acoustical Society of America*, *133*(4), 2055–2068. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/133/4/10.1121/1.4792355 doi: http://dx.doi.org/ 10.1121/1.4792355

Mannerheim, P. V. H. (2008). *Visually Adaptive Virtual Sound Imaging Using Loudspeakers*, Unpublished doctoral dissertation, University of Southampton, Southampton, UK.

Moore, A. H., Tew, A. I., & Nicol, R. (2010). An initial validation of individualized crosstalk cancellation filters for binaural perceptual experiments. *Journal of Audio Engineering Society*, *58*(1-2), 36–45. Retrieved from www.aes.org/ e-lib/browse.cfm?elib=15240

Morse, P. M., & Ingard, K. U. (1986). *Theoretical Acoustics* (pp. 306–312). Princeton, NJ: Princeton University Press.

Nelson, P. A., & Elliott, S. J. (1993). *Active Control of Sound*. London, UK: Academic Press.

Nelson, P., Kirkeby, O., Takeuchi, T., & Hamada, H. (1997). Sound fields for the production of virtual acoustic images. *Journal of Sound and Vibration*, *204*(2), 386–396. Retrieved from www.sciencedirect.com/ science/article/pii/S0022460X97909676 doi: http://dx.doi.org/ 10.1006/jsvi.1997.0967

Nelson, P. A., & Rose, J. F. W. (2005). Errors in two-point sound reproduction. *The Journal of the Acoustical Society of America*, *118*(1), 193–204. Retrieved from http://scitation.aip.org/content/asa/ *BIBLIOGRAPHY*73journal/jasa/118/1/10.1121/1.1928787 doi: http://dx.doi.org/ 10.1121/1.1928787

Nicol, R. (2010). *Binaural Technology* (pp. 30–44). New York, NY: Audio Engineering Society Inc.

Papadopoulos, T., & Nelson, P. A. (2010). Choice of inverse filter design parameters in virtual acoustic imaging systems. *Journal of Audio Engineering Society*, *58*(1-2), 22–35. Retrieved from www.aes.org/e-lib/browse.cfm ?elib=15239

Parodi, Y. L., & Rubak, P. (2010). Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers. *The Journal of the Acoustical Society of America*, *128*(3), 1045–1055. Retrieved from http://scitation.aip.org/content/asa/ journal/jasa/128/3/10.1121/1.3467763 doi: http://dx.doi.org/ 10.1121/1.3467763

Parodi, Y. L., & Rubak, P. (2011a). Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations. *Journal of Audio Engineering Society*, *59*(5), 304–320. Retrieved from www.aes.org/ e-lib/browse.cfm?elib=15931

Parodi, Y. L., & Rubak, P. (2011b). A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers. *Journal of Audio Engineering Society*, *59*(7-8), 487–497. Retrieved from www.aes.org/ e-lib/browse.cfm?elib=15974

Sæbø, A. (2001). *Influence of Reflections on Crosstalk Cancelled Playback of Binaural Sound*, Unpublished doctoral dissertation, Norwegian University of Science and Technology, Trondheim, Norway.

SreenivasaRao, C., Mahalakshmi, N., & VenkataRao, D. (2012). Real-time dsp implementation of audio crosstalk cancellation using mixed uniform partitioned convolution. *Signal Processing: An International Journal (SPIJ)*, *6*(4), 118–127. Retrieved from www.scribd.com/ document/299653040/Real-time-DSP-Implementation-of -Audio-Crosstalk-Cancellation-using-Mixed-Uniform -Partitioned-Convolution

Takeuchi, T., & Nelson, P. A. (2002). Optimal source distribution for binaural synthesis over loudspeakers. *The Journal of the Acoustical Society of America*, *112*(6), 2786–2797. Retrieved from http://scitation.aip.org/ content/asa/journal/jasa/112/6/10.1121/1.1513363 doi: http://dx.doi.org/10.1121/1.1513363

Takeuchi, T., & Nelson, P. A. (2007). Subjective and objective evaluation of the optimal source distribution for virtual acoustic imaging. *Journal of Audio Engineering Society*, *55*(11), 981–997. Retrieved from www.aes.org/e-lib/ browse.cfm?elib=14181

Takeuchi, T., Nelson, P. A., & Hamada, H. (2001). Robustness to head misalignment of virtual sound imaging systems. *The Journal of the Acoustical Society of America*, *109*(3), 958–971. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/ 109/3/10.1121/1.1349539 doi: http://dx.doi.org/10.1121/1.1349539

Ward, D. B. (2001). On the performance of acoustic crosstalk cancellation in a reverberant environment. *The Journal of the Acoustical Society of America*, *110*(2), 1195–1198. Retrieved from http://scitation.aip.org/ content/asa/journal/jasa/110/2/10.1121/1.1386635 doi: http://dx.doi.org/10.1121/1.1386635

Ward, D. B., & Elko, G. (1999, May). Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. *Signal Processing Letters, IEEE*, *6*(5), 106–108. doi: 10.1109/97.755428

Xie, B. (2013). *Head-Related Transfer Function and Virtual Auditory Display* (2nd ed., pp. 283–326). Plantation, FL: J. Ross Publishing.

Yang, J., Gan, W.-S., & Tan, S.-E. (2003). Improved sound separation using three loudspeakers. *Acoustics Research Letters Online*, *4*(2), 47–52. Retrieved from http://scitation.aip.org/content/asa/ journal/arlo/4/2/10.1121/1.1566419 doi: http://dx.doi.org/10 .1121/1.1566419

Appendix A

Derivation of the Optimal XTC Filter

Here we carry out the derivation of Equations (5.73)–(5.75) following the approach outlined in the section "Impulse Response."

We start by factoring the expressions appearing in Equations (5.69) and (5.70), which, we note, have the same denominator, into the following products of terms:

$$H^{[O]}_{LL_{\mathrm{I,II}}}(i\omega) = H^{[O]}_{RR_{\mathrm{I,II}}}(i\omega) = (\psi_0 + \gamma\psi_1)\psi_a, \tag{A1}$$

$$H^{[O]}_{LR_{\mathrm{I,II}}}(i\omega) = H^{[O]}_{LR_{\mathrm{I,II}}}(i\omega) = (\mp\psi_0 + g\gamma e^{i\omega\tau_c}\psi_1)\psi_a, \tag{A2}$$

where

$$\psi_0 = \gamma^2\left[\pm x - g^2(1 + e^{2i\omega\tau_c})\right], \tag{A3}$$

$$\psi_2 = \sqrt{g^2 \mp x + 1}, \tag{A4}$$

$$\psi_a = \frac{1}{(g^2 \mp x + 1) \pm 2\gamma x\sqrt{g^2 \mp x + 1}}. \tag{A5}$$

The term Ψ_a can be factored as

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

where

$$\psi_2 = \frac{1}{2\gamma x}, \tag{A6}$$

$$\Psi_3 = \frac{1}{\sqrt{g^2 \mp x + 1}} \tag{A7}$$

$$\Psi_4 = 2\gamma x, \tag{A8}$$

$$\Psi_5 = \frac{1}{8\gamma^3 x^3}, \tag{A9}$$

$$\Psi_6(c) = \frac{1}{1 - cx^{-1}}, \tag{A10}$$

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 Equations (A1) and (A2) becomes:

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

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

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 Equation (A3) can be readily found by substituting back $2g\cos(\omega\tau_c)$ for x and carrying out the IFT integration:

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

- Ψ_1: Making the substitution $b \equiv g^2 + 1$ in Equation (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

$$\binom{k}{m} = \begin{cases} \dfrac{k!}{m!(k-m)!} & \text{if } 0 \le m \le k, \\ 0 & \text{if } m < 0 \text{ or } k < m. \end{cases}$$

Since $0 < g < 1$, we have $|x| = 2g|\cos(\omega\tau_c)| < g^2 + 1 = b$, and the series in Equation (A18) always converges. However, as $g \to 1$, $b \to 2$, and when $\omega\tau_c \to 2n\varpi$ 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

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

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

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

the IFT of Ψ_1 can be expressed as

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

- Ψ_2: Explicitly, Equation (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

$$\sec(\omega\tau_c) = \frac{1}{\sqrt{1-\sin^2(\omega\tau_c)}}, \tag{A22}$$

$$\text{if } 2n\pi - \frac{\pi}{2} < \omega\tau_c < 2n\pi + \frac{\pi}{2}$$

with $n = 0,1,2,3,4, \ldots$

Furthermore, we note that since

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

the arguments of the inverse cosine function in Equation (5.59) obeys the condition:

$$0 < \frac{(g^2+1)\gamma^2 - 1}{2g\gamma^2} \le 1 \tag{A24}$$

which leads us to write

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

In light of this expression and Equation (5.55), we conclude that the conditions for the validity of Equation (A22) are always satisfied in Branch-I bands.

Similarly, we find that $\sec(\omega\tau_c)$ can be expressed as $-1\Big/\sqrt{1-\sin^2(\omega\tau_c)}$ 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 $\omega\tau_c = (2n+1)\pi/2$, with $n = 0,1,2,3,4, \ldots$. From the band division conditions in Equations (5.55) and (5.57) 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 Equation (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

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

Missing pages for copyright reasons.

and g), we first set $y(c) = +1$ and $y(c) = -1$, and solve for $\eta^{+}(c)$ and $\eta^{-}(c)$, respectively, to find, for Branch-I bands,

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

$$\eta^{-}(c_2) = \cos^{-1}\left(\frac{f(g,\gamma)-1}{16g\gamma^2}\right), \tag{A49}$$

and, for Branch-II bands,

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

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

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

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

Using these four explicit expressions, along with the definition of φ given by Equation (5.59), we find that the inequalities in Equations (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 Equations (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 both in Branch-I and in Branch-II bands explicitly in terms of g and γ:

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

This convergence condition is illustrated in the region plot of Figure 5.A.1, where the black-shaded 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 $g \simeq 1$.

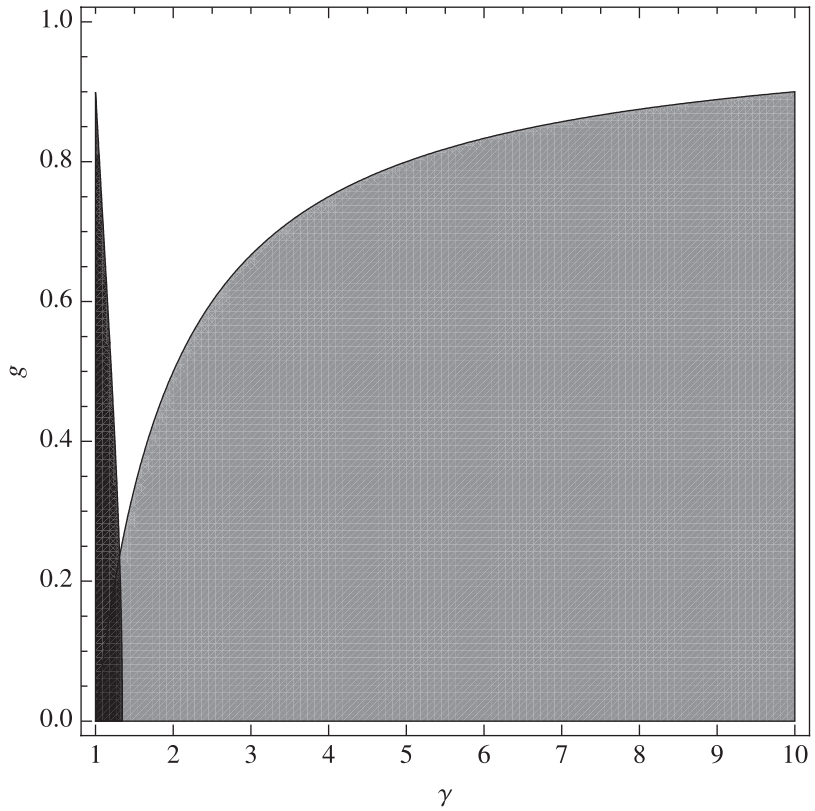

Figure 5.A.1 Region plot showing the allowed values for g and γ (white). The black-shaded region is where the series convergence condition in Equation (A54) is not satisfied, and the grey-shaded region is where the general condition in Equation (5.51) is violated.

Aside from the series convergence condition above, γ must satisfy the general condition given by Equation (5.51) (whose region of violation is shaded in grey in Figure 5.A.1). Therefore, we combine both conditions in the following expression:

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

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

Now that we have found the convergence condition for the series in Equation (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 Equation (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 Equation (A29), and the associated IFT in Equation (A30), we arrive at the sought expression for the IFT of $\Psi_6(c)$:

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

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

Appendix B

Numerical Verification

The optimal XTC IRs derived in the previous appendix were evaluated for the typical case of $g = 0.985$ and $\Gamma = 7$ dB, and plotted in Figure 5.8. To verify the validity of the IRs and assess the effect of the number of terms in the series expansions, we calculated their Fourier transforms and compared the resulting spectra to those obtained from the frequency-domain expressions of the section "Frequency Response." An example is shown in Figure 5.B.1 for the Branch-I part of the XTC spectrum (top panel) and that of the envelope spectrum (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 IRs, with the exception of ψ_1 and ψ_3, which, due to their slow convergence at and near the frequencies $\omega\tau_c = 2n\varpi$ with $n = 0, 1, 2, 3, 4, \ldots$, 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 Figure 5.B.1), 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 1,000 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 IRs is not shown, as the resulting spectra exhibit the same behavior as that described above.

Acknowledgments

The author wishes to thank Joseph Tylka for his help in checking the manuscript and updating the citations, and J. S. Bach for his Mass in B Minor, whose reproduction in 3D was a main motive for this work.

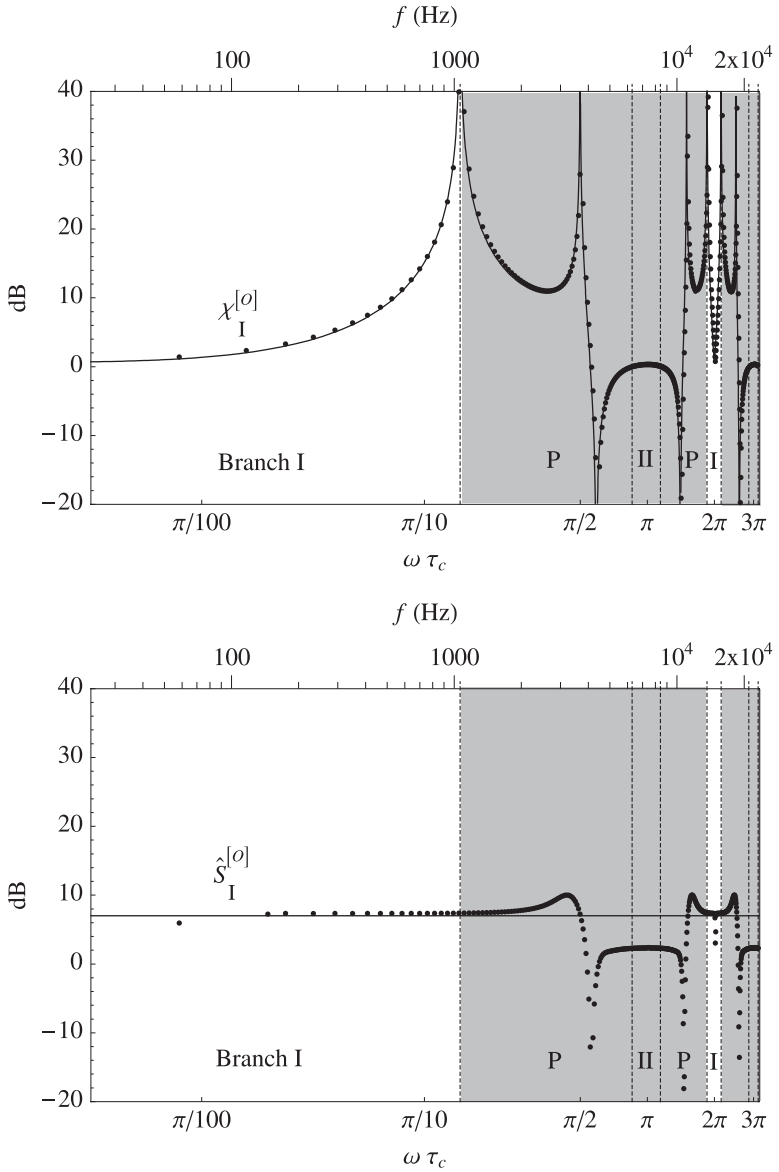

Figure 5.B.1 XTC spectrum of optimal filter for Branch-I bands, $\chi_I^{[O]}(\omega)$, shown in top panel, and the associated envelope spectrum, $\hat{S}_I^{[O]}(\omega)$, shown in bottom panel. The small dots represent the spectra calculated by taking the Fourier transform of the Branch-I part of the IRs derived in Appendix A. (The IRs are shown graphically in the top panel of Figure 5.8.) 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 2,500 terms were used. The hard curve in the top panel is the Branch-I XTC spectrum calculated directly from Equation (5.63), and the horizontal line in the bottom panel is the Branch-I envelope spectrum $\hat{S}_I^{[O]}(\omega)$, with Γ = 7 dB. (Other parameters are the same as for Figure 5.2.) 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.)