Audio Engineering Society
Convention Paper 10614
Presented at the 153rd Convention
2022 October



# Multiband time-domain crosstalk cancellation

Alberto Vancheri[1], Tiziano Leidi[1], Thierry Heeb[1], Loris Grossi[1], and Noah Spagnoli[1]

[1] *University of Applied Sciences and Arts of Southern Switzerland*

Correspondence should be addressed to Tiziano Leidi (tiziano.leidi@supsi.ch)

**ABSTRACT**

Pioneered in the sixties, Crosstalk Cancellation (CTC) allows for immersive sound reproduction from a limited number of loudspeakers. Upcoming virtual reality and augmented reality applications, as well as widespread availability of 3D audio content, have boosted interest in CTC technologies over the recent years. In this paper, we present a novel multiband approach to CTC, evolving and superseding our original work based on modeling of the system's geometrical acoustics. This new solution, whilst keeping a simple processing model, offers improved CTC effectiveness, reduced residual coloration and wider bandwidth. The enhanced performance of our new approach has been confirmed by laboratory experiments.

## 1 Introduction

Crosstalk Cancellation (CTC) is an audio processing technique allowing to deliver immersive 3D audio by controlling the signals received at listener's ears from a set of speakers. In particular, CTC may be used in such a way that contributions from a given source speaker are received at one of the listener's ears and cancelled out at the other. Using binaural encoded content and two source speakers, this allows for proper rendering of binaural cues at the listener's ears, hence providing an immersive 3D audio experience.

CTC has been an active field of research for many years and many approaches have been proposed among which the Recursive Ambiophonic Crosstalk Elimination (RACE) initially presented by Glasgal [1] is a time-domain processing approach that inspired many modern solutions. In our previous work [2], we introduced a variation of the RACE algorithm based on the modelling and control of the propagation of acoustical waves from the sources to the listener's ears. Presenting high similarity to the ray tracing approach used in computer graphics rendering, it provides increased robustness by regularization of the order of the cancellation signals cues.

Based on the same principles, we present a novel solution featuring a multiband approach to crosstalk cancellation. The introduction of frequency-band dependent CTC allows us to obtain enhanced cancellation, reduced coloration and wider bandwidth, whilst, at the same time, achieving system robustness similar to the solution presented in our previous paper. Sections 4 to 6 provide detailed information on the theoretical and practical aspects of the newly proposed solution. Finally, in section 7, we present results from laboratory experiments performed on a real-world implementation of the newly proposed multiband approach to CTC. Results show that the new system clearly outperforms the solution presented in our previous work.

## 2 CTC overview

Binaural reproduction over loudspeakers using CTC has been an active research topic for a long time and im-

mersive 3D audio content as well as upcoming virtual and augmented reality applications have lately boosted interest in the field. The principle of CTC was pioneered by Bauer [3] in the early sixties whereas the first patent in the field was filed by Atal et al. in 1966 [4] and commercial applications appeared about 20 years later (Cooper Bauck Transaural). The works of Masiero et al. [5] and Gardner [6] provide a good overview of CTC technologies.

A CTC system based on two loudspeakers can be described by the following $z$-domain matrix equation:

$$E(z) = H(z)S(z) \tag{1}$$

where $E(z) = (E_1(z), E_2(z))^T$ represents the left and right ears signals, $S(z) = (S_1(z), S_2(z))^T$ the left and right speaker signals and $H(z)$ is a 2x2 matrix whose coefficients $h_{ij}$ are the transfer functions from speaker $i$ to ear $j$.

By introducing a filter at the input of the system, represented by a matrix $CTC(z)$, the transfer function results in:

$$E(z) = H(z)CTC(z)S(z) \tag{2}$$

Perfect crosstalk cancellation is achieved if

$$E(z) = kz^{-\delta}S(z) \tag{3}$$

where $k$ is a gain factor and $z^{-\delta}$ is pure delay. In other words, $CTC(z)$ is an approximation of the inverse of the forward path matrix $H(z)$, up to the gain factor $k$, combined with a delay for causality reasons. Computation of $CTC(z)$ is thus closely related to a matrix inversion problem and is generally ill-defined due to the typically non-minimum phase nature of $H(z)$. The same formalism can obviously be applied to systems with more than two loudspeakers or for multiple users.

It has been shown by Parodi [7] that correct sound source localization requires cancellation levels of 20 dB and more. According to Choueiri [8], very high levels of boost (reaching 30 dB and above) may be required at frequencies where the matrix inversion is problematic. At such frequencies, approximation errors can result in high deviations between expected and computed values, hence reducing crosstalk cancellation effectiveness. These stringent constraints usually result in a small sweet spot (i.e. the area where the crosstalk cancellation is effective) and require $CTC(z)$ to be adapted to the actual listener position as shown by Lee and Lee [9]. A variation of the RACE algorithm able to support non-central user positions has been presented by Cecchi et al. [10]. Whilst this paper presents some similarities with our own previous work [2], our solution, based on cancellation complexes, uses truncated impulse responses instead of a full recursive scheme, which provide benefits in terms of system stability.

Many approaches to achieve robust, artefacts-free CTC have been studied, among which: optimized loudspeaker positions (and types) as suggested by Ward and Elko [11] and Takeuchi and Nelson [12], mapping the inversion problem to an $L_\infty$ minimization problem as proposed by Rao et al. [13] or stochastic approaches based on random perturbation matrices as studied by Xu et al. [14]. Technologies inspired by sound field reproduction such as the analytical spectral division method studied by Qiao [15] have also been successfully applied to the CTC problem.

Frequency-dependent or multiband CTC has been identified as an efficient mean to enhance CTC robustness and effectiveness. For instance, in his work [8], Choueiri presents a method for designing optimal CTC filters for two loudspeakers systems based on frequency-domain regularization, where different frequency bands are associated with different analytically derived CTC impulse responses. Speaker arrays and beam-forming are also widely used approaches for the implementation of frequency-dependent CTC. Solutions based on combinations of different technologies, each being optimized for a given frequency band have also been researched as exemplified by the works of Ma et al. [16] or Bruschi et al. [17]. These works present multiband CTC solutions based on a combination of beam-forming for mid/high frequencies and modified versions of the RACE algorithm for low frequencies.

Compared to the cited approaches, the novel multiband CTC solution presented in this paper is a RACE-inspired, full time-domain approach based on a simple, unified geometrical acoustics model, that offers both high cancellation effectiveness and robustness as has been confirmed by simulations and laboratory experiments.

Furthermore, if the listener is allowed to move freely (as is the case in virtual reality applications), $H(z)$ becomes time variant and, consequently, the filter $CTC(z)$

has to be updated in real-time to track the user's position and orientation. A significant part of recent research in the field of crosstalk cancellation has been centered on the real-time computation, optimization and smooth regularization of the inverse approximation $CTC(z)$ of the time-varying system forward path $H(z)$. Our approach being full time-domain based, provides natural support for time-varying CTC filters as they can be updated on a sample by sample basis. However, in this paper, we will not cover the topic of moving users which is a subject of on-going research by our team.

## 3 Background of our approach

Our approach to crosstalk cancellation is based on the notion of cancellation complex, a system made of three sources $S_0$, $S_1$ and $S_2$ and two receivers $E_1$ and $E_2$ (the ears of the user), as depicted in figure 1. The ear $E_1$ is called the target ear. The speaker $S_0$ outputs a signal $x_0(t)$. The sources $S_1$ and $S_2$ work in such a way that the signal $y_1(t)$ received at the target ear is the propagation of $x_0(t)$ from $S_0$ to $E_1$, whereas the signal received at the non target ear $E_2$ is $y_2(t) = 0$. Sources $S_1$ and $S_2$ provide the recursive cancellation signals needed to attain this goal. In this paper, $x_i(t)$ will indicate a signal emitted from source $S_i$ and $y_j(t)$ a signal received at the ear $E_j$.

In order to compute the impulse responses $m_1(t)$ and $m_2(t)$ associated to speakers $S_1$ and $S_2$, we consider six acoustic paths labelled each with a couple of indexes $(i, j)$, where $i$ refers to the source and assumes values in $\{0,1,2\}$ and $j$ refers to the ears and assumes values in $\{1,2\}$. For instance the path $(0,2)$ refers to the acoustic path between source $S_0$ and non target ear $E_2$.

The solution presented in our previous work [2] was based on the assumption that the sound propagation along the path $(i, j)$ amounts to a global gain $g_{ij}$ and a global delay $\tau_{ij}$ applied to the signal emitted from the source $i$: if $x_i(t)$ is the signal emitted from the source $i$, then the signal received at the ear $j$ will be $y_j(t) = g_{ij}x_i(t - \tau_{ij})$.

This propagation law is described by the impulse response

$$h_{ij}(t) = g_{ij}\delta(t - \tau_{ij}) \tag{4}$$

or, equivalently, by the transfer function

$$H_{ij}(\omega) = g_{ij}\exp(-i\omega\tau_{ij}) \tag{5}$$

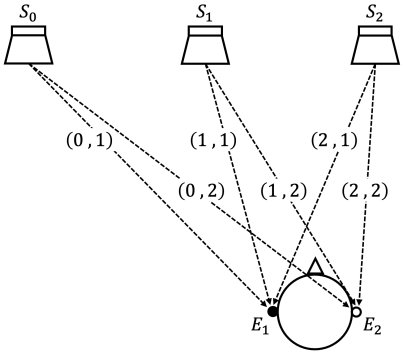



**Fig. 1:** Structure of a cancellation complex. The acoustic paths between the speaker $S_i$ and the ear $E_j$ is labelled with $(i, j)$. The original sound is emitted from the speaker $S_0$ and is directed to the target ear $E_1$ along the path $(0,1)$. The speakers $S_1$ and $S_2$ cooperate to cancel the crosstalk generated along the path $(0,2)$.

The delay $\tau_{ij}$ and the gain $g_{ij}$ include contributions from free propagation in space and from the scattering of the sound wave on the head. More precisely, the delay is written as a sum $\tau_{ij} = \tau_i^{(0)} + \tau_{ij}^{(H)}$ of two contributions, where $\tau_i^{(0)}$ is due to free propagation in space and $\tau_{ij}^{(H)}$ represents the contribution of the head. The propagation term is simply given by $\tau_i^{(0)} = \frac{L_i}{c}$, where $L_i$ is the distance between the source and the center of the head and $c$ is the speed of sound. A similar definition can be given for $g_{ij}$, which is the product of a propagation gain $g_i^{(0)}$, which is dependent on the distance $L_i$ and a head related contribution $g_{ij}^{(H)}$.

With these assumptions and notations, it can be shown that, for a static user, the impulse response $m_j(t)$, $j = 1,2$, associated to the speaker $S_j$ is an infinite sequence of delayed and dumped pulses:

$$m_j(t) = \sum_{p=1}^{\infty} m_{j,p}\delta(t - t_{j,p}) \tag{6}$$

where $m_{j,p}$ and $t_{j,p}$ are, by definition, the magnitudes and the releasing times of the $p$-th pulse. The pulses $m_{j,p}\delta(t - t_{j,p})$ are the $p$-th order response from the speaker $S_j$. The impulse responses are truncated at a given cancellation order $N$.

It can be shown that the releasing times $t_{j,p}$ are arithmetic sequences:

$$t_{j,p} = t_{j,1} + (p-1)T \tag{7}$$

where $t_{1,1} = \tau_{02} - \tau_{22} + \tau_{21} - \tau_{11}$, $t_{2,1} = \tau_{02} - \tau_{22}$ and $T = \tau_{12} + \tau_{21} - \tau_{11} - \tau_{22}$.

In a similar way, it is easy to show that the magnitudes of the pulses $m_{j,p}$ are given by geometric sequences:

$$m_{j,p} = m_{j,1} G^{p-1} \tag{8}$$

where $m_{1,1} = \frac{g_{02}g_{21}}{g_{11}g_{22}}$, $m_{2,1} = \frac{g_{02}}{g_{22}}$ and $G = \frac{g_{12}g_{21}}{g_{11}g_{22}}$.

The $p$-th order response from the speaker $S_2$ will be called the $p$-th order cancellation signal, whereas the $p$-th order response from $S_1$ will be called the $p$-th order decoloration signal. The reason for this terminology is that the $p$-th component of $m_2(t)$ emitted from $S_2$ is aimed at cancelling a crosstalk received at the non-target ear $E_2$, whereas the $p$-th component of $m_1(t)$ emitted from $S_1$ is aimed at cancelling the coloration induced at the target ear $E_1$ by the crosstalk produced by $m_2(t)$. Here the use of the terms "cancellation" and "decoloration" is slightly improper: we conventionally call "cancellation signal" a signal aimed at cancelling a crosstalk at the non-target ear and "decoloration signal" a signal aimed at cancelling a crosstalk at the target ear. But it is well known that a loss of spatiality of the acoustic scene and the coloration effects induced by crosstalk cannot be separated in a so simple way.

If $X_0(\omega)$ is, in frequency domain, the input signal to be reproduced, the CTC system described above does not properly map $X_0(\omega)$ to the target ear $E_1$ up to a global gain and delay as in the usual formulation of the CTC problem. The signal received at the target ear will indeed be $Y_1(\omega) = P_{01}(\omega)X_0(\omega)$, where $P_{01}(\omega)$ is the transfer function from the speaker $S_0$ to the target ear $E_1$, inclusive of the listener's Head Related Transfer Function (HRTF). In order to obtain $X_0(\omega)$ one has to pre-process $X_0(\omega)$ in order to compensate the effect of the HRTF.

Each audio signal to be reproduced needs a different cancellation complex but the same speaker can be used in different complexes and also be used twice in the same complex. For example, in ordinary setups for the reproduction of two audio channels, only two speakers $A$ and $B$ can be used, with the first complex assigning the speakers $S_0 = A$, $S_1 = A$ and $S_2 = B$, and the second complex the speakers $S_0 = B$, $S_1 = B$ and $S_2 = A$. The only constraints are that the same speaker cannot be used in both roles $S_0$ and $S_2$ within the same complex (indeed, $S_2$ provides the first order cancellation of the cross-talk produced by $S_0$), and a complex cannot assign $S_1$ and $S_2$ to the same speaker.

In experimental results shown in our previous work [2], we computed the delays and gains (at a reference frequency of 2000 Hz) for an elliptic head using a model inspired by Brown and Duda [18]. This approach, based on an approximation of the HRTF with a single gain and a single delay, has proven to provide good cancellation performances but in a narrow frequency band (performances of single band CTC are also presented in section 7).

## 4 Multiband time-domain CTC

With the solution described in section 3, the cancellation effectiveness is good only in a narrow frequency band. This is not surprising as unique values of gain $g$ and delay $\tau$ correspond to an HRTF with constant magnitude and linear phase response, which is not the case in reality. Such an approximation is expected to be effective only in a small interval around a frequency $\omega_0$ where the delay and gain of the HRTF are $\tau$ and $g$ respectively. This suggests that a generalisation of the approach, where several values of gain and delays at different frequencies are considered, can lead to a consistent improvement of CTC effectiveness. In this section we will introduce such a generalisation.

The more general version of CTC presented in this section is referred to as a multiband time-domain approach because both the design of the CTC system, which is based on a generalisation of impulse responses equation 6, and the implementation of the corresponding algorithm are rooted in the time-domain (frequency-domain methods are used in the next sections only as tools for analysis). The term multiband refers to the general design of the CTC system, which is based on time-domain CTC subsystems of the type described by equation 6, each dedicated to the processing of the input signal in a specific frequency band. The whole frequency range, from 0 Hz up to the Nyquist frequency, is subdivided into $n$ disjoint frequency bands $B^{(1)}, B^{(2)}, ..., B^{(n)}$. The bands $B^{(1)}$ and $B^{(n)}$ are called boundary bands and cover low and high frequency ranges where we will not apply crosstalk cancellation. Border bands $B^{(1)}$ and

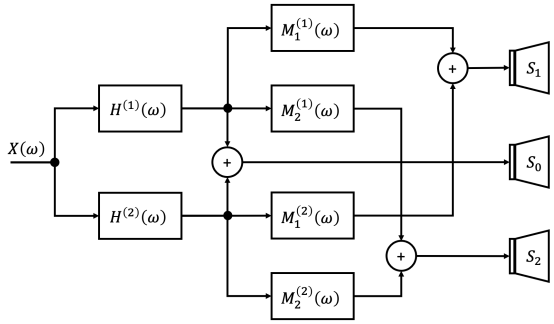



**Fig. 2:** Circuital scheme of the multiband CTC system. For simplicity of the representation, only two bands have been considered.

$B^{(n)}$ are associated with low pass and high pass filters respectively, whereas all other bands are associated with band pass filters. The frequency response and the related impulse response of the filter associated to the band $B^{(k)}$ will be designated by $H^{(k)}(\omega)$ and $h^{(k)}(t)$ respectively.

The circuital schema of the multiband CTC system is shown in figure 2. The system is made of $n$ subcomponents, one for each band $B^{(k)}$, with associated impulse responses $m_j^{(k)}(t)$, and related frequency responses $M_j^{(k)}(\omega)$, defined as in equation 6 with band-specific releasing times and magnitudes $t_{i,p}^{(k)}$ and $m_{j,p}^{(k)}$:

$$m_j^{(k)}(t) = \sum_{p=1}^{\infty} m_{j,p}^{(k)} \delta(t - t_{j,p}^{(k)}) \tag{9}$$

The releasing times $t_{j,p}^{(k)}$ and the magnitudes $m_{j,p}^{(k)}$ are defined as in equations 7 and 8, with the band-specific delays and gains $\tau_{ij}^{(k)}$ and $g_{ij}^{(k)}$. The band-specific impulse responses in equation 9 will be called partial impulse responses.

For simplicity, we will treat boundary bands as normal cancellation bands with gains $g_{ij}^{(1)} = g_{ij}^{(n)} = 0$ in such a way that no cancellation signal is sent in these bands. The original signal $X(\omega)$ to be reproduced is filtered with the filter system described above and each component $H^{(k)}(\omega)X(\omega)$ is passed to the corresponding subsystem. Each subsystem computes the partial cancellation signals to be emitted from the cancellation speakers $S_1$ and $S_2$ using band specific impulse responses defined as in equation 9. The overall cancellation signal is obtained by adding up the partial responses. As will be explained in section 5, we use IIR filters $H^{(k)}(\omega)$ for band separation and the signal emitted from the speaker $S_0$ is not the input signal $X(\omega)$ of the system, but the sum of the bands signals $X_0(\omega) = \sum_{k=1}^{n} H^{(k)}(\omega)X(\omega)$.

As stated at the end of section 3, the CTC system described above maps the signal emitted from the speaker $S_0$ to the target ear $E_1$ by means of the transfer function $P_{01}(\omega)$ connecting $S_0$ with $E_1$. Depending on the concrete application, this can be a virtue or a shortcoming. If needed, the band signals generated by the filters $H^{(k)}(\omega)$ can be processed with the delays and gains $\tau_{01}^{(k)}$ and $g_{01}^{(k)}$ before recombining them in the signal emitted from the speaker $S_0$, in such a way that the effect of the HRTF is compensated and the target ear $E_1$ receives (an approximation of) the input signal.

## 5 Band filters

In this section, we present the practical implementation of the band filters used in the multiband approach to CTC presented in this paper. As explained in the previous section, the whole frequency range from 0 Hz to half the sampling rate is subdivided into $n$ non-overlapping bands $B^{(1)}, B^{(2)}, ..., B^{(n)}$, each associated to specific delay and gain values $\tau_{ij}^{(k)}$ and $g_{ij}^{(k)}$ with $k = 1, ..., n$. As the frequency range of interest for CTC is usually of bandpass nature, no cancellation is applied for bands $B^{(1)}$ and $B^{(n)}$, which is equivalent to setting $g_{ij}^{(1)} = 0$ and $g_{ij}^{(n)} = 0$.

Each band $B^{(1)}, B^{(2)}, ..., B^{(n)}$ is defined by a low cutoff frequency $f_{low}^{(k)}$ and a high cutoff frequency $f_{high}^{(k)}$ with $f_{high}^{(k)} = f_{low}^{(k+1)}$, except for band $B^{(1)}$ which only has a high-cutoff frequency and band $B^{(n)}$ which only has a low-cutoff frequency.

The filters used for band separation are 4$^{\text{th}}$ order IIR filters configured as Linkwitz-Riley cross-overs between bands. This choice is motivated by sufficient out-of-band attenuation and by the fact that the gain of the sum of the low-pass and high-pass branches of a Linkwitz-Riley crossover amounts to 0 dB across the whole spectrum. In other words, the sum of the low-pass and high-pass branches behaves like an all-pass filter, having a flat amplitude response with a smoothly changing phase response. Additional 2$^{\text{nd}}$ order all-pass filters

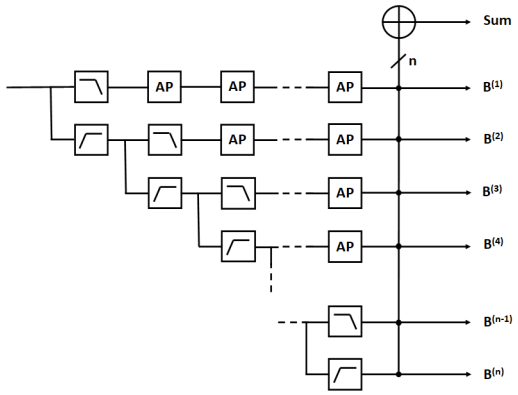



**Fig. 3:** Band filters structure.

are used to maintain a coherent phase response across all bands. The resulting filter structure is illustrated in figure 3.

In our approach to CTC, the original signal is emitted by the source speaker $S_0$. However, as the bands extraction crossovers introduce phase shifts, the same shifts must be applied to the original signal for proper CTC operation. This can be achieved by running the input signal through a set of all-pass filters corresponding to the different crossovers. Alternatively, the outputs of bands $B^{(1)}$ to $B^{(n)}$ can be summed together to construct a phase shifted original signal. Both approaches have been studied but the latter was preferred as it makes sure that the original signal is exactly equal to the sum of the band signals, thus reducing potential error sources for CTC.

FIR-based band filters could also have been used, but computational complexity is generally higher, especially for low cutoff frequencies. A further advantage of the IIR-based approach is that filter coefficients are computed from simple, closed-form formulas. For instance, this could allow for dynamic adaptation of crossover frequencies when possible onsets of CTC instability are detected.

## 6 Theoretical analysis of the multiband approach

In this section, we will define the cancellation residual and the cancellation effectiveness and derive formula for measuring these quantities at the first cancellation order in a multiband approach. The cancellation residual $r(t)$ is defined as the sum $r(t) = y_2^{(ct)}(t) + y_2^{(c)}(t)$ of the cross talk $y_2^{(ct)}(t)$ associated to the direct sound emitted from $S_0$ and received at the non target ear $E_2$ and the corresponding cancellation signal $y_2^{(c)}(t)$. If cancellation is perfect, the residual $r(t)$ is 0. The cancellation effectiveness $C(t_1,t_2)$ in a time window between $t_1$ and $t_2$ is the ratio of the energy $E[r(t)]$ of the residual $r(t)$ over the energy $E\left[y_2^{(ct)}(t)\right]$ of the crosstalk $y_2^{(ct)}(t)$:

$$C(t_1,t_2) = \frac{E[r(t)]}{E\left[y_2^{(ct)}(t)\right]} \tag{10}$$

In this section, the cancellation effectiveness will be computed in frequency domain with the impulse responses truncated in such a way that $m_{1,p}^{(k)} = 0$ for all $p$ and $m_{2,p}^{(k)} \neq 0$ only for $p = 1$ (see equation 9). This quantity will be called first order effectiveness. We will indicate with $P_{ij}(\omega)$ the true transfer function from the source $S_i$ to the ear $E_j$, inclusive of the free propagation in space and the HRTF of the listener. Let us consider an input $X(\omega)$. As written in section 4, the signal emitted from the speaker $S_0$ is $X_0(\omega) = \sum_{k=1}^{n} H^{(k)}(\omega)X(\omega)$. Hence, the cross talk produced at the non-target ear $E_2$ is

$$Y_2^{(ct)}(\omega) = P_{02}(\omega) \sum_{k=1}^{n} H^{(k)}(\omega)X(\omega) \tag{11}$$

The component of the system associated to the band $B^{(k)}$ receives as input the filtered signal $H^{(k)}(\omega)X(\omega)$ and computes the first order cancellation signal applying a forecast factor $F^{(k)}(\omega)$ to this input to obtain the partial signal $F^{(k)}(\omega)H^{(k)}(\omega)X(\omega)$, where:

$$F^{(k)}(\omega) = \frac{g_{02}^{(k)}}{g_{22}^{(k)}} \exp\left(i\omega\left(\tau_{22}^{(k)} - \tau_{02}^{(k)}\right)\right) \tag{12}$$

The forecast factor is used by the subsystem $k$ to forecast the intensity and timing of the cancellation signal: it takes into account a forward propagation in time $\tau_{02}^{(k)}$ from $S_0$ to $E_2$ and a back propagation in time $\tau_{22}^{(k)}$ from $E_2$ back to to $S_2$ and similar for the gains.

The partial responses of all the subsystems are then added together, the sign is changed, and the resulting

signal is emitted from $S_2$ and propagated with $P_{22}(\omega)$ to $E_2$. The resulting cancellation signal is:

$$Y_2^{(c)}(\omega) = -P_{22}(\omega)\sum_{k=1}^{n} F^{(k)}(\omega)H^{(k)}(\omega)X(\omega) \quad (13)$$

The cancellation residual is the sum of the crosstalk $Y_2^{(ct)}(\omega)$ (eq. 11) and the cancellation signal $Y_2^{(c)}(\omega)$ (eq. 13):

$$R(\omega) = Y_2^{(ct)}(\omega) + Y_2^{(c)}(\omega) \quad (14)$$

Finally, the relative complex amplitude $A(\omega) = \frac{R(\omega)}{X(\omega)}$ of the residual $R(\omega)$ can be written in the following way:

$$A(\omega) = \sum_{k=1}^{n} H^{(k)}(\omega)\left(P_{02}(\omega) - P_{22}(\omega)F^{(k)}(\omega)\right) \quad (15)$$

The first order cancellation effectiveness $C(\omega)$ is the ratio of the powers of the residual $R(\omega)$ (eq. 14) over the crosstalk $Y_2^{(ct)}(\omega)$ (eq. 11):

$$C(\omega) = 20\log_{10}\left(\left|\frac{R(\omega)}{Y_2^{(ct)}(\omega)}\right|\right) \quad (16)$$

Following the same approach, it is also possible to compute the first order coloration residual defined as the sum of the crosstalk generated by the first order cancellation signal emitted from $S_2$ and received at $E_1$, and the corresponding cancellation signal emitted from $S_1$. The definition of the cancellation and coloration residuals include all the set of true propagation functions $P_{ij}(\omega)$ and the whole series of gains and delays $g_{ij}^{(k)}$ and $\tau_{ij}^{(k)}$. Cancellation and coloration residuals up to a given order $N$ can also be computed in a similar way.

The minimisation of the $L^2$ norm of the residual relative amplitude $A(\omega)$ is a complex optimisation process that involves the choice of the bands $B^{(k)}$, of the associated filters $H^{(k)}(\omega)$ and of the delay and gain $\tau_{ij}^{(k)}$ and $g_{ij}^{(k)}$. The problem becomes even more complicated if we think that the optimisation should take into account different head positions and that a system of filters which is well adapted to a given position could perform less optimally in another one. This optimization problem is a current research activity within our group. To validate the newly proposed multiband approach, we will limit ourselves here to a choice of filters based on general considerations and use gain and delays derived from experimental measures of the propagation functions $P_{ij}(\omega)$ at the center of each band.

A short inspection of the forecast factors $F^{(k)}(\omega)$ (eq. 12) shows that they approximate a mapping from $P_{22}(\omega)$ to $P_{02}(\omega)$ and that the residual is small when this mapping is accurate: $F^{(k)}(\omega) \simeq \frac{P_{02}(\omega)}{P_{22}(\omega)}$.

When a propagation model $P_{ij}(\omega)$ is available (for instance from experimental measures) these criteria give us a simple method for the choice of the values of the gains $g_{ij}^{(k)}$ and delays $\tau_{ij}^{(k)}$ based on the minimisation of the error $P_{ij}(\omega) - g_{ij}^{(k)}\exp\left(-i\omega\tau_{ij}^{(k)}\right)$ within the band $B^{(k)}$. This is not the optimal solution because of the overlapping between the transfer functions of the filters.

Finally, it is worthwhile to spend some words about the role of the overlapping of transition bands of the filters in the crosstalk cancellation process. Let us consider a sinusoidal input $x(t) = \exp(i\omega t)$ and the related output $x_0(t) = \sum_{k=1}^{n} H^{(k)}(\omega)\exp(i\omega t)$ from the speaker $S_0$. This sinusoidal input will induce, in a cancellation speaker $S_j$, $j = 1,2$, a sinusoidal response $x_j(t)$ at order $p$:

$$x_j(t) = \sum_{k=1}^{n} H^{(k)}(\omega)m_{j,p}^{(k)}\exp\left(-i\omega t_{j,p}^{(k)}\right)\exp(i\omega t) \quad (17)$$

where $m_{j,p}^{(k)}$ and $t_{j,p}^{(k)}$ are the delay and gain at order $p$ in the sub-component $k$ defined as in equations 7 and 8 for the component $k$. The transfer function connecting $x_0(t)$ with $x_j(t)$ is

$$H(\omega) = \frac{\sum_{k=1}^{n} H^{(k)}(\omega)m_{j,p}^{(k)}\exp\left(-i\omega t_{j,p}^{(k)}\right)}{\sum_{k=1}^{n} H^{(k)}(\omega)} \quad (18)$$

The form of this transfer function suggests that the phase delay and gain relating $x_0(t)$ with $x_j(t)$ is a sort of weighted average of the gains and delays $m_{j,p}^{(k)}$ and $t_{j,p}^{(k)}$ applied by the single components. This is true especially when the overlapping between filters is not negligible. We consider this behaviour as an advantage, as overlapping filters realize a sort of smooth interpolation of gains and delays associated to the bands (see figure 4). On the other hand, a system of almost ideal

filters would be equivalent to a step approximation of the transfer functions $P_{ij}(\omega)$. The use of overlapping filters enables a much smoother approximation.

This suggests a way to optimize the configuration of the multiband CTC system. Following the method for the construction of filters exposed in section 5, a CTC multiband system is defined by the following data:

- a cancellation range $[f_{low}, f_{high}]$, that is a frequency range where cancellation is applied;
- the partition of the cancellation range $[f_{low}, f_{high}]$ into $n-2$ bands with $n > 3$;
- values of band gains and delays $\tau_{ij}^{(k)}$ and $g_{ij}^{(k)}$ along each of the six paths $(i, j)$ for $k = 2, .., n-2$

The optimal CTC system can be selected by minimizing the difference between the interpolated gain and delays (the solid black lines in figure 4) and the corresponding quantities defined using the transfer functions $P_{ij}(\omega)$ derived from laboratory measures with a cost function depending on the number of bands $n$ to avoid solutions based on a too large number of bands.

## 7 Experimental Results

Laboratory experiments to assess the performance of the new multiband approach to CTC have been conducted using the same setup and conditions as in our previous work [2]. The experimental setup consists in a pair of stand mounted loudspeakers, located 2 m in front of the user and spaced by 60 cm. Measurements are made using a dummy head equipped with in-ear microphones. The input solicitation signal is a Gaussian-modulated sinusoidal pulse centered at 6 kHz with a relative bandwidth of 2.

As introduced in section 6, the contribution of the head to the delays $\tau_{ij}^{(k)}$ and gains $g_{ij}^{(k)}$ have been computed based on recordings performed, in laboratory, by means of the same dummy head. The system operates at a sampling rate of $F_s = 48$ kHz and the selected frequency bands have been set according to Table 1. CTC is applied on bands $B^{(2)}$, $B^{(3)}$ and $B^{(4)}$. The emission signal consists in the sum of all band signals.

The performance of the system has been analysed using two criteria: the cancellation effectiveness defined in equation 10 and the residual coloration defined

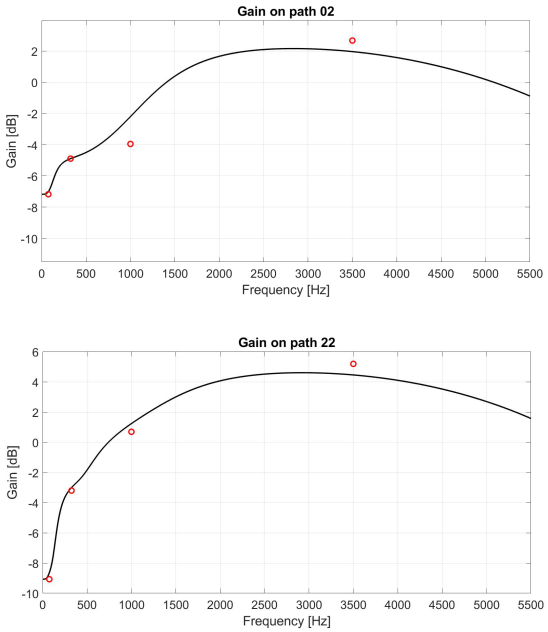



**Fig. 4:** Smoothing (solid black lines) induced by the band filters on the band coefficients $g_{02}^{(k)}$ and $g_{22}^{(k)}$ (red circles) for $k = 2, 3, 4$

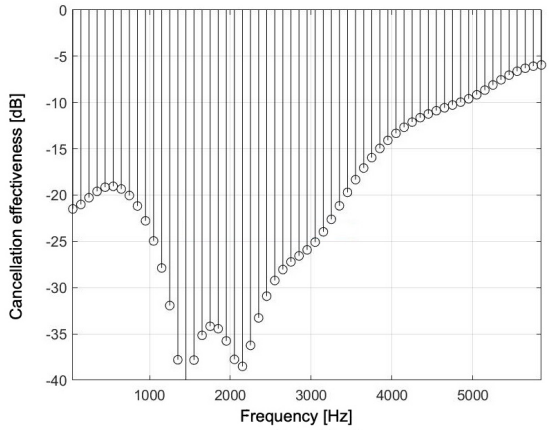
as $Col(t_1, t_2) = \frac{E[c(t)]}{E\left[y_1^{(0)}(t)\right]}$, where $c(t) = y_1^{(0)}(t) - y_1(t)$. $y_1(t)$ and $y_1^{(0)}(t)$ are the signals received at the target ear when the cancellation is on and off respectively.

The cancellation effectiveness measured on the dummy head's left ear is depicted in figure 5a, for a multiband CTC with cancellation up to order $N = 7$. The spectrum of the residual coloration for the same multiband CTC experiment is represented in figure 6a.

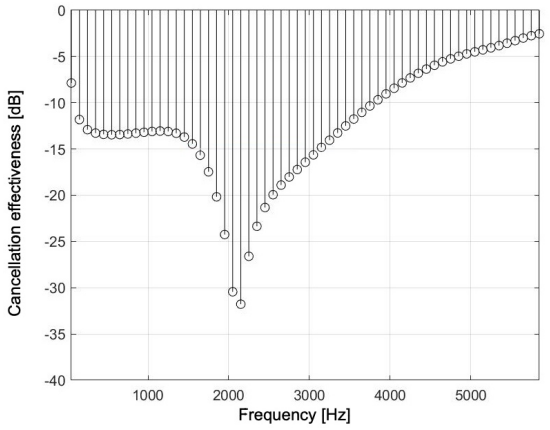
As a comparison, figure 5b and figure 6b depict the cancellation effectiveness and the residual coloration

| Band | $f_{low}$ [Hz] | $f_{high}$ [Hz] |
|---|---|---|
| $B^{(1)}$ | (0) | 150 |
| $B^{(2)}$ | 150 | 500 |
| $B^{(3)}$ | 500 | 1500 |
| $B^{(4)}$ | 1500 | 5500 |
| $B^{(5)}$ | 5500 | $(F_s/2)$ |

**Table 1:** Experimental setup frequency bands.

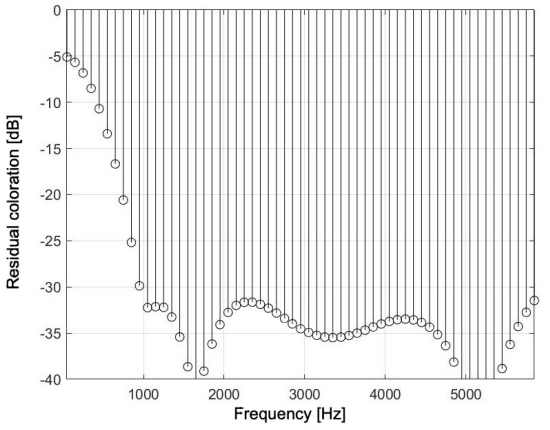



**(a)** Multiband CTC.

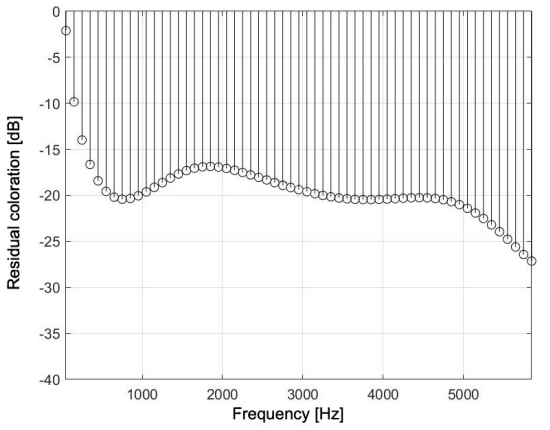



**(b)** Monoband CTC.

**Fig. 5:** Cancellation effectiveness (in dB) with impulse responses truncated at order 7.




**(a)** Multiband CTC.




**(b)** Monoband CTC.

**Fig. 6:** Residual coloration (in dB) with impulse responses truncated at order 7.

obtained for a single band approach. The difference between the monoband and multiband approaches is in the band where the CTC is applied. In the monoband approach, the frequency band of the used IIR filter is from 0 Hz to 4500 Hz. Please note that, compared to the approach described in [2], the type of filter has changed from FIR to IIR.

By comparing the plots, it can be noticed that the multiband approach allows obtaining better performances than the monoband on the entire band where the CTC is applied. By using frequency dependent head contributions for gains and delays, it is possible to overcome the limitations described in our previous paper [2]. Results could further be optimized by splitting $B^{(4)}$ (1500 Hz to 5500 Hz) into two or more bands to better take into account the fact that HRTFs generally don't exhibit constant group delay nor flat magnitude response over the frequency range of $B^{(4)}$.

## 8 Summary and conclusions

In this paper, we have presented a new time-domain approach to CTC, based on a multiband approximation of the acoustical propagation model from the sound sources to the listener's ears. The proposed solution extends the concept of cancellation complexes, introduced in our previous work [2], to multiple frequency bands allowing for a better approximation of the HRTF related part of the propagation. Laboratory experiments have been conducted, confirming the enhanced performance of the multiband approach, especially in terms

of CTC effectiveness, reduced coloration and achievable bandwidth of interest.

These encouraging results motivate further developments of the proposed multiband, time-domain approach to CTC. Future research directions will focus on integration of dynamic user tracking and support for arbitrary user positions and orientations, especially through optimization of the residuals.

This work is an extension of an initial research program funded by Innosuisse, the Swiss funding agencies for innovative technologies, under grant 42471.1 IP-ICT INXS-3D.

## References

[1] Glasgal, R., "360 Localization via 4.x RACE Processing," *Audio Engineering Society 123rd Convention*, 2007.

[2] Vancheri, A., Leidi, T., Heeb, T., Grossi, L., Spagnoli, N., and Weiss, D., "Geometrical Acoustics Approach to Crosstalk Cancellation," in *Proceedings of the Audio Engineering Society 152nd Convention*, 2022.

[3] Bauer, B. B., "Stereophonic Earphones and Binaural Loudspeakers," *J. Audio Eng. Soc*, 9(2), pp. 148–151, 1961.

[4] Atal, B. S. and Schroeder, M. R., "Apparent sound source translator," 1966, uS Patent 3,236,949.

[5] Masiero, B. S., Fels, J., and Vorländer, M., "Review of the crosstalk cancellation filter technique," 2011.

[6] Gardner, W. G., "3-D Audio Using Loudspeakers," 1998.

[7] Lacouture Parodi, Y., *"A systematic study of binaural reproduction systems through loudspeakers: A multiple stereo-dipole approach"*, Ph.D. thesis, 2010.

[8] Choueiri, E. Y., "Optimal Crosstalk Cancellation for Binaural Audio with Two Loudspeakers," Self-published, 2010.

[9] Lee, K.-S. and Lee, S.-P., "A real-time audio system for adjusting the sweet spot to the listener's position," *IEEE Transactions on Consumer Electronics*, 56, 2010.

[10] Cecchi, S., Primavera, A., Virgulti, M., Bettarelli, F., Li, J., and Piazza, F., "An efficient implementation of acoustic crosstalk cancellation for 3D audio rendering," in *2014 IEEE ChinaSIP*, pp. 212–216, 2014, doi:10.1109/ChinaSIP.2014.6889234.

[11] Ward, D. and Elko, G., "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," *IEEE Signal Processing Letters*, 6(5), pp. 106–108, 1999, doi:10.1109/97.755428.

[12] Takeuchi, T. and Nelson, P. A., "Optimal source distribution for binaural synthesis over loudspeakers," *The Journal of the Acoustical Society of America*, 112(6), pp. 2786–2797, 2002, doi:10.1121/1.1513363.

[13] Rao, H. I. K., Mathews, V. J., and Park, Y.-C., "A Minimax Approach for the Joint Design of Acoustic Crosstalk Cancellation Filters," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), pp. 2287–2298, 2007, doi:10.1109/TASL.2007.905149.

[14] Xu, H., Wang, Q., Xia, R., Li, J., and Yan, Y., "A Stochastic Approximation Method with Enhanced Robustness for Crosstalk Cancellation," *Chinese Journal of Electronics*, 26(6), pp. 1269–1275, 2017, doi:https://doi.org/10.1049/cje.2017.09.035.

[15] Qiao, Y. and Choueiri, E., "Real-time Implementation of the Spectral Division Method for Binaural Personal Audio Delivery with Head Tracking," in *Proceedings of the Audio Engineering Society 151st Convention*, 2021.

[16] Ma, X., Hohnerlein, C., and Ahrens, J., "Concept and Perceptual Validation of Listener-Position Adaptive Superdirective Crosstalk Cancellation Using a Linear Loudspeaker Array," *J. Audio Eng. Soc*, 67(11), pp. 871–881, 2019.

[17] Bruschi, V., Cecchi, S., Bruschi, V., Ortolani, N., and Piazza, F., "Immersive sound reproduction in real environments using a linear loudspeaker array," 2019.

[18] Brown, C. and Duda, R., "An efficient HRTF model for 3-D sound," in *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 4 pp.–, 1997, doi:10.1109/ASPAA.1997.625596.