Audio Engineering Society

Convention Paper 10548

Presented at the 152nd Convention
2022 May, In-Person and Online

Geometrical Acoustics Approach to Crosstalk Cancellation

Alberto Vancheri[1], Tiziano Leidi[1], Thierry Heeb[1], Loris Grossi[1], Noah Spagnoli[1], and Daniel Weiss[2]

[1] *University of Applied Sciences and Arts of Southern Switzerland*
[2] *Weiss Engineering Ltd*

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

ABSTRACT

Crosstalk Cancellation (CTC) is a signal processing technique allowing for immersive sound reproduction from a limited number of loudspeakers. Pioneered in the sixties, CTC has lately gained much attraction due to upcoming Augmented Reality and Virtual Reality applications and generalization of 3D audio content. In this paper, we present a novel time-domain approach to CTC based on modeling of the system's geometrical acoustics. Our solution provides a simple processing model, as well as means to address robustness issues and adaptation to arbitrary listener positions.

1 Introduction

Crosstalk Cancellation (CTC) is a sound reproduction technique that has been explored since the sixties. Its primary goal is to provide control over the sound reproduced at a listener's ears by a set of loudspeakers. Considering a pair of speakers, CTC may be used in such a way that contributions from the left speaker are cancelled out at the listener's right ear and contributions from the right speaker are cancelled out at the listener's left ear, providing a direct 1:1 mapping from audio channels to listener's ears. With binaural encoded source material, this will result in the binaural cues being properly reproduced at listener's ears and hence providing full 3D immersion from a pair of loudspeakers.

As will be detailed in section 2, many approaches for CTC have been proposed. Among the many proposed CTC solutions, Recursive Ambiophonic Crosstalk Elimination (RACE), initally presented by Glasgal [1], is a recursive time-domain digital processing technique that inspired many modern solutions. Interestingly, recursive processing in the analog domain was proposed as early as 1978 in a patent by Iwahara and Mori [2].

In this paper, we present an extension and generalization of the RACE principle, able to efficiently handle some of the implementation challenges of real-time CTC. Our approach is inspired by geometrical acoustics and provides a different perspective on CTC able to simplify the design and implementation of flexible solutions that adapt to different situations.

Laboratory experiments performed on a concrete implementation of our approach demonstrated a cancellation effectiveness of more than 20 dB, whilst limiting CTC side effects and minimizing reconfiguration impact (in particular, for modifications of the listener's position). In addition, this paper provides a method to assess CTC robustness as a function of loudspeakers and user positions, which helps in the mitigation of some of the difficulties arising during design, development and configuration of real-world CTC systems.

2 CTC principle and solutions

The principle of CTC was pioneered by Bauer [3] in the early sixties. The first patent in the domain was filed by Atal et al. in 1966 [4] and commercial applications started to appear about 20 years later (Cooper Bauck Transaural). Since then, CTC has been a very active field in both academic and commercial research, with a recent surge in activity due to progress in the fields of Augmented Reality and Virtual Reality. An overview of CTC approaches and technologies can be found in the works of Masiero et al. [5] and Gardner [6].

A CTC system based on two loudspeakers can be described by the following z-domain equations where h_{11} is the transfer function from the left speaker to the left ear, h_{12} is the transfer function from the left speaker to the right ear and similarly for h_{21} and h_{22}:

$$E_1(z) = h_{11}(z)S_1(z) + h_{12}(z)S_2(z) \quad (1)$$

$$E_2(z) = h_{21}(z)S_1(z) + h_{22}(z)S_2(z) \quad (2)$$

where $E_1(z)$ and $E_2(z)$ are the left and right ears signals and $S_1(z)$ and $S_2(z)$ are the left and right speaker signals. Eq. 1 and Eq. 2 can be rewritten in matrix form as:

$$E(z) = H(z)S(z) \quad (3)$$

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

$$E(z) = H(z)S(z)CTC(z) \quad (4)$$

Perfect crosstalk cancellation is achieved if

$$E(z) = kz^{-\delta}S(z) \quad (5)$$

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. This formalism can be extended 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, especially in the case of front to back disambiguation. Clearly, if the listener moves, $H(z)$ and, consequently, $CTC(z)$ will be time variant.

Computation of $CTC(z)$ is hence closely related to matrix inversion and is generally ill-defined due to the typically non-minimum phase nature of the system $H(z)$. According to Choueri [8], very high levels of boost (reaching 30dB and above) are required at frequencies where the inversion is problematic. At such frequencies, small errors in the computation of the inverse approximation can result in high deviations between expected and computed values due to the high gain. Such deviations will inevitably result in reduced crosstalk cancellation effectiveness. Most recent research activities in the field of crosstalk cancellation have been centered on the computation, optimization and regularization of the inverse approximation $CTC(z)$ of the system forward path $H(z)$.

Furthermore, in the case of a moving listener, $CTC(z)$ has to be updated in real-time to track modifications in system geometry, resulting in possibly high computational load. Both time-domain and frequency-domain approaches have been studied, with the latter gaining increased attention due to the computational load reduction of frequency-domain processing. Kirkeby et al. [9] have demonstrated that $CTC(z)$ can be written as an infinite sum of geometrically decaying delayed impulses, meaning that optimal CTC impulse responses are infinitely long, hence correspond to IIR filters. This may result in aliasing and non-causality when using the Discrete Fourier Transform for frequency domain approaches due to the DFT's periodicity. Whilst these effects can be alleviated by extending the frequency resolution of the matrix $H(z)$ or by applying time-domain windowing prior to domain transformation, artifact-free transitions from one frame's $CTC(z)$ to the next present a major challenge in the case of a moving user. Novel, highly efficient sliding Fourier Transform algorithms such as the one presented by Park in [10] may also contribute to artifacts reduction for frequency domain approaches. Time-domain approaches, on the other hand, do not suffer from these drawbacks as $CTC(z)$ can easily be updated on a sample by sample basis. The novel CTC solution presented in this paper follows a time-domain approach, inspired by RACE and its variations such as the work by Cecchi et al. [11].

It is important to notice that CTC performance is also highly influenced by physical constraints. These include sensitivity to listener's movements and to placement of physical loudspeakers as presented by Bai and Lee [12], as well as limited spatial region of CTC effectiveness (sweet spot) studied by Kirkeby et al. [13]. Recent research has been addressing these points and various solutions have been proposed. For instance,

Yang et al. [14] and Nawfal et al. [15] use multiple loudspeakers to extend the sweet spot, whereas Qiao and Choueri [16] use multiple loudspeakers to provide CTC at different spatial locations. Robustness to listener's movements can be improved by methods such as the sum and difference approach suggested by Kim et al. [17]. Dynamic sweet spot adjustment to listener's position based on user tracking with motion sensors is commonly used today as presented by Lentz and Schmitz [18].

Computation of $CTC(z)$ can also be interpreted as the solution to an L_∞ minimization problem as shown by Rao et al. [19]. The main drawback of this approach is its high computational load which makes it unsuitable for real-time applications. Finally, in [8], Choueri presents a method for designing optimal CTC filters for two loudspeakers systems based on frequency-dependent regularization, where different frequency bands are associated with different analytically derived CTC impulse responses. This multi-bands approach results in significantly improved regularization of the crosstalk cancellation process.

3 RACE as a starting point for time-domain crosstalk cancellation

In this section, we briefly recall the idea of time domain CTC implemented by RACE. A simple (and approximate) way to consider CTC in time domain is to figure out the sound propagation from a couple of loudspeakers S_1 and S_2 to the ears E_1 and E_2 of the user as the propagation of single pulses. If the loudspeaker S_1 emits a unit pulse p_1 directed to the ear E_1 at time $t = 0$, then a crosstalk will reach the ear E_2 at a later time τ_{12} with an amplitude g_{12}. In this paper, we will indicate with τ_{ij} and g_{ij} the propagation delays and gains along the path from S_i to E_j. In order to cancel the crosstalk, we need to emit from the source S_2 a pulse p_2 of amplitude $-g_{12}/g_{22}$ at time $\tau_{12} - \tau_{22}$. This simple description implies that the cancellation procedure is recursive: the pulse p_2 used to cancel the crosstalk generated by p_1 at E_2 will in turn produce crosstalk at E_1 that needs to be cancelled with a pulse p_3 emitted from E_1 and so on.

RACE implements this simple idea by means of a recursive circuit based on delay lines and multiplicative blocks. In our approach, instead of using a recursive circuit, we directly work with the impulse responses of the system in continuous time. As will be shown later, this enables a flexible generalisation to more complex architectures and a clear treatment of issues related to the stability and robustness of the cancellation process. Our approach enables us to truncate recursion at a given order through windowing of the impulse responses in accordance with an optimisation criterion based on the metrics defined in this paper. Indeed, in theory, cancellation is perfect only when full recursion is active, but in non-optimal situations it may be convenient to limit the recursion order to reduce the amount of energy emitted from the loudspeakers.

A problem that can be addressed in a very natural way in our approach concerns implementations with a user moving in space. The delays and gains needed to achieve a good cancellation effectiveness depend in general on the position of the user in space and on the rotation angles of the head. If the user is in motion, the length of the delay lines and the values of the gains have to be updated in principle at each audio sample. Furthermore, the delay and gain values to be used at a given time might differ from the values associated with the static user position, as the radial velocities of the ears with respect to the sound sources have to be taken into account, especially if such velocities are high. In our approach, the kinematics of pulses can be easily generalized to the case of moving users, avoiding the re-sampling procedures needed to manage time varying delay lines. This enables a natural and simple approach to such situations.

A further problem to which our approach offers a natural solution is related to the causality of the impulse responses. The standard RACE implementation works only in a limited set of configurations because in many cases the cancellation process requires impulse responses with an anticausal part. Anticausal filters are implementable in the standard RACE architecture if some system latency is allowed. A simple way to manage anticausal components in RACE consists in adding additional delay lines on the external paths of the circuit. Working instead directly with impulse responses does not require any special trick to manage such configurations because anticausal components appear naturally in the computation of the impulse responses.

4 Cancellation complexes and their relations with RACE

In this section, we will introduce the notion of cancellation complex, the basic building block of our approach,

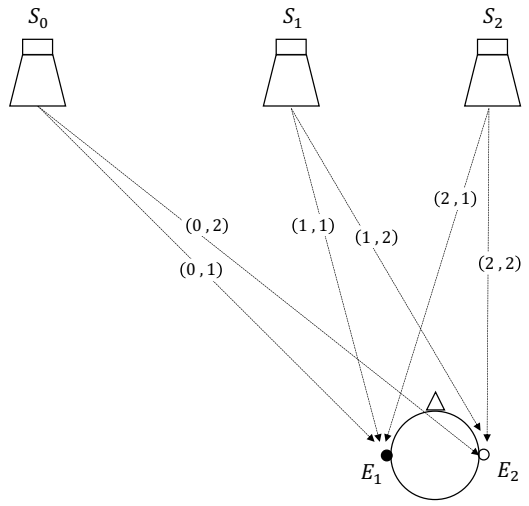

Fig. 1: Structure of a cancellation complex. The acoustic paths between the speaker S_i and the ear E_j is labelled with (i, j). The original sound is emitted from the speaker S_0 and is directed to the target ear E_1 along the path $(0,1)$. The speakers S_1 and S_2 cooperate to cancel the crosstalk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

and introduce the related notation. Next, we will derive the impulse responses associated with a cancellation complex.

As illustrated in figure 1, a cancellation complex is a structure made of three sound sources S_0, S_1 and S_2, and a target ear. The cancellation complex works in such a way that the signal emitted from the speaker S_0 is received at the target ear after propagation through space whereas nothing is received at the other ear (the non-target ear). In order to attain this goal, we consider six acoustic paths labelled with a double index (i, j), where $i \in \{0,1,2\}$ and $j \in \{1,2\}$. The first index i refers to the speaker, whereas the second index j refers to the ear, $j = 1$ to the target ear and $j = 2$ to the non-target ear. For instance the acoustic path $(1,2)$ is the path from the speaker S_1 to the non-target ear.

In usual implementations, the speaker S_0 coincides physically with the speaker S_1. Furthermore, the same physical speaker can be used in more than one complex. A constraint to the distributions of these roles to the speakers is that S_0 can never coincide with S_2. Further constraints are related to the stability of the complex as a linear system and will be addressed at the end of this section. To implement a system that is equivalent to a standard RACE with a user located in front of two speakers A and B, we use two complexes: complex 1, with speakers (A,A,B) and left target ear and complex 2 with speakers (B,B,A) and right target ear.

In this way, the number of audio-channels and corresponding settings can be flexibly increased to build complicated setups with limited impact on the complexity of the implementation. Cancellation complexes have to be considered as building blocks for complex setups. In implementations with more than two loudspeakers, it is possible to use several complexes sharing the same speaker S_0 to manage CTC for one audio channel: the cancellation process can be distributed among these complexes in such a way that optimisation criteria for robustness are met. To avoid artifacts, it is convenient to use a fixed speaker S_0 for the emission of the original sound channel and switch between different couples of cancellation speakers S_1 and S_2, as the user changes position.

The acoustics of the complex is described through the six propagation delays τ_{ij} and the six propagation gains g_{ij} (see section 3). The cancellation complex can be interpreted, for a static user, as a linear time invariant system with one input and two outputs. In order to compute the impulse responses of the complex we consider a unit pulse delivered from S_0 at the time $t = 0$. The task of the complex is to cancel the crosstalk generated along the path $(0,2)$ in such a way that, at the target ear, only the direct sound propagating along the path $(0,1)$ is received. The impulse response can be described by means of four sequences $t_n^{(1)}, m_n^{(1)}, t_n^{(2)}, m_n^{(2)}$ with $n \in \mathbb{N}$: the speaker S_1 emits the n-th pulse with intensity $m_n^{(1)}$ at time $t_n^{(1)}$ and a similar interpretation holds for the speaker S_2. In our implementation, all these data are computed and stored as real numbers in floating point format and converted to the sampling rate in use only immediately before the signal is sent to the loudspeakers. This enables us to process the data in a flexible way when the user is moving, i.e. when the system is no longer linear time invariant.

With these settings and notations, it can be shown that, for a static user, the impulse response $m^{(j)}(t)$, $j = 1,2$, associated to the speakers S_1 and S_2 are, in a continuum time formulation, $m^{(j)}(t) = \sum_{n=1}^{\infty} m_n^{(j)} \delta(t - t_n^{(j)})$ where $m_n^{(j)}$ and $t_n^{(j)}$ are, by definition, the magnitudes and the

releasing times of the pulses. It can be shown that the releasing times $t_n^{(1)}$ and $t_n^{(2)}$ are arithmetic sequences:

$$t_n^{(1)} = t_{in}^{(1)} + (n-1)T \tag{6}$$

$$t_n^{(2)} = t_{in}^{(2)} + (n-1)T \tag{7}$$

where $t_{in}^{(1)} = \tau_{02} - \tau_{22} + \tau_{21} - \tau_{11}$, $t_{in}^{(2)} = \tau_{02} - \tau_{22}$ 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 $m_n^{(1)}$ and $m_n^{(2)}$ are given by geometric sequences:

$$m_n^{(1)} = m_{in}^{(1)} G^{n-1} \tag{8}$$

$$m_n^{(2)} = -m_{in}^{(2)} G^{n-1} \tag{9}$$

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

The delays τ_{ij} and the gains g_{ij} include contributions from free propagation in space and from the head related transfer function (HRTF). More precisely, τ_{ij} is the sum of two contributions $\tau_{ij}^{(0)} = \frac{l_i}{c}$, where l_i is the distance between the source S_i and the center of the head, c is the speed of sound, and $\tau_{ij}^{(H)}$ represents a head related contribution: $\tau_{ij} = \tau_{ij}^{(0)} + \tau_{ij}^{(H)}$. A similar definition can be given for g_{ij}, which is the product of a propagation gain and a head related gain. Modelling the contribution of the head in this way is equivalent to the approximation of the HRTF with $h_{ij}(\omega) = g_{ij}\exp(-i\omega\tau_{ij})$.

The quantities G and T play a crucial role in the feasibility of the cancellation process and in the analysis of the robustness. First of all, the complex is stable if and only if $G < 1$, as is clear from the equations for the magnitudes written above. As G approaches the value $G = 1$, the pulses decay more and more slowly with time and this leads to a larger and larger amount of energy emitted from the loudspeakers. Furthermore, it is easy to see from the equations above that a sinusoidal wave emitted from the speaker S_0 induces the emission of a superposition of sinusoidal waves at the same frequency from S_1 and S_2. These sinusoidal waves are delayed by T one from the other and their amplitudes scale as a power of $\frac{1}{G}$. It is clear that at frequencies, which are multiples of $f_0 = \frac{1}{T}$, the interference is constructive, whereas at frequencies given by odd multiples of $\frac{1}{2T}$ the interference is destructive. In real conditions, this process leads to coloration of the sound emitted from the speakers S_1 and S_2, and perception of sound spatiality could be compromised if cancellation at the ears is not sufficient. Thus, configurations where G is near to 1 and $f_0 = \frac{1}{T}$ is inside the frequency band where cancellation is performed, may lack robustness and require a regularisation procedure. We will address this topic in the next section.

5 Windowed cancellation complexes

The infinite impulse responses derived in the previous section can be truncated at any order N. It has proven to be useful to introduce two different methods of truncation named counterlateral and ipsilateral. In the N-th order counterlateral truncation, both the impulse responses $m^{(1)}(t)$ and $m^{(2)}(t)$ are truncated at the order N. In this way, the speaker S_2 does not send the signal that should cancel the crosstalk produced by the N-th pulse emitted by the speaker S_1 at the non target ear (counterlateral with respect to the target ear). In the N-th order ipsilateral cancellation, the impulse response $m^{(1)}(t)$ is truncated at order $N-1$, whereas the impulse response $m^{(2)}(t)$ is truncated at the order N. In this way, the speaker S_1 does not send the signal that should cancel the crosstalk produced by the N-th pulse emitted by the speaker S_2 at the target ear. The ipsilateral and counterlateral truncation schemes might lead to a significant amount of non cancelled energy at the target ear and at the non target ear respectively, with different potential impacts on acoustic scene spatiality and sound coloration.

It is easy to measure coloration and missed cancellation using the analytical form of the impulses responses given in the previous section. In case of ipsilateral cancellation at the order N, the amplitude of crosstalk at the target ear (the not cancelled residual) is $g_{21}m_n^{(2)}$. The amplitude at the target ear of the sound coming from the speaker S_0 is g_{01} (recall that the pulse emitted from S_0 has amplitude 1). If we define the missed cancellation as the ratio of the non cancelled amplitude over the amplitude of the original sound at the target ear, we obtain the following measure for the coloration error in ipsilateral truncation at order N:

$$E_I(N) = \frac{g_{02}g_{21}}{g_{01}g_{22}} G^{N-1} \tag{10}$$

We see that the error scales as a power of G. In experimental situations, we have found that a value of G around 0.5 is favorable for good CTC.

The error in counterlateral truncation is defined as the ratio of the amplitude of the not cancelled crosstalk at the non-target ear $m_n^{(1)} g_{12}$ over the amplitude g_{02} of the crosstalk produced by the signal originally emitted from the speaker S_0. Hence, the measure for the error associated to counterlateral truncation is:

$$E_C(N) = G^N \tag{11}$$

The explicit form of the impulse responses enables us to estimate the coloration effect of the unstable frequency $f_0 = \frac{1}{T}$ when a truncation at order N is applied. As discussed early in this paper, sinusoidal waves emitted from the loudspeakers S_1 and S_2 at a frequency f_0 will interfere constructively. It can be shown that the resulting amplitude is in the order of $\frac{1}{1-G}$ when no truncation is applied. This value becomes very high as G approaches unity. When truncation at order N is applied, the resulting amplitude will be in the order of $\frac{1-G^N}{1-G}$. For $G = 0.9$ we obtain an amplification by a factor 10 without truncation that reduces to $\frac{1-0.9^5}{1-0.9} \approx 4.7$ if a truncation at order $N = 6$ is applied. On the other hand, reducing the cancellation order increases the ipsilateral or the counterlateral errors $E_I(N)$ and $E_C(N)$ defined above. Thus, an optimal value of N has to be chosen based on these metrics.

Configurations with high values of G are far from unusual, especially when the user is located in a side position or the head is rotated. A simple strategy to deal with these challenging configurations consists in introducing a tolerance threshold $G_{th} < 1$: when $G < G_{th}$ the impulse response is truncated to a preset order N, which in standard configurations typically lies between 5 and 10. In configurations where $G_{th} \leq G < 1$, the following operations are undertaken: (1) the cancellation order is set to an optimized value N_0, (2) a number n_0 with $1 \leq n_0 \leq N_0$ is selected, (3) a windowing of the impulse response is applied to the components $m_n^{(1)}$ and $m_n^{(2)}$ of the impulse responses for $n_0 \leq n \leq N_0$ (this windowing can simply consist in substituting G with G_{th}) and (4) a truncation schema, either ipsilateral or counterlateral, is selected. All these parameters can be optimized when an optimisation criterion based on the cancellation and coloration metrics is considered.

In situations where $G > 1$ the system is unstable and therefore exact cancellation is not feasible. Nevertheless, compromises are possible, for example by reducing the cancellation order to $N = 1$. With an ipsilateral truncation scheme we obtain coloration at the target ear that can be managed with a decoloration filter. An alternative corrective measure would be to moderate G by limiting its value below 1.

Fig. 2: Setup used for laboratory experiments.

6 Experimental Results

Laboratory experiments assessing the performance of the proposed CTC approach have been conducted in a regular room with a real-time personal-computer based implementation on a setup composed of a pair of loudspeakers and an in-ear microphones equipped dummy head (see figure 2). The experimental setup has been initially configured by positioning the loudspeakers in front of the listener at a distance of 2 meters. The distance between the loudspeakers has been set at 60 cm.

As input solicitation signal a Gaussian-modulated sinusoidal pulse with a center frequency of 6000 Hz and a fractional bandwidth of 2 has been used. With these parameters, the pulse has rich spectral content in the band ranging from 0 to 12000 Hz and a duration in the order of 0.2 ms. In all the performed experiments, the cancellation signals have been generated by initially

low-pass filtering the solicitation signal to limit its spectral content to the frequency band that is normally of interest for CTC: 0 Hz to 4500 Hz.

The head related contributions to the delays τ_{ij} and gains g_{ij} have been computed using a standard head model, according to the work of Brown and Duda [20], with a reference frequency f_0 of 2000 Hz. We choose the reference frequency $f_0 = 2000Hz$ because there is a sufficiently large frequency band around this value which is relevant for crosstalk cancellation and where the errors on gains and delays remain sufficiently low. The validity of this choice is confirmed by the results presented below in this section.

We have measured the performance of the system by means of two quantities: the cancellation effectiveness and the residual coloration. Before defining them, we point out that the words "coloration" and "cancellation" are used here with a somehow improper meaning. We refer the words "coloration" and "missed cancellation" to a loss of performance at the target and non-target ear respectively for simplicity, even if coloration and spatiality are influenced by the performances at both ears. Let us consider the signals $x_1(t)$ and $x_2(t)$ received at the target and non-target ear respectively when the cancellation procedure is on. Let us define $x_1^{(0)}(t)$ and $x_2^{(0)}(t)$ as the corresponding signals measured when the cancellation procedure is off. If we indicate with $E[x(t)]$ the energy of a signal $x(t)$, we define the cancellation effectiveness and the residual coloration respectively as the energy ratios $\frac{E[x_2(t)]}{E[x_2^{(0)}(t)]}$ and $\frac{E[x_1(t)]}{E[x_1^{(0)}(t)]}$ measured in dB.

Figure 3 provides the spectrum of the cancellation effectiveness measured on one of the dummy head ears, for a CTC with cancellation order $N = 7$ and counterlateral truncation scheme. Figure 4 provides the spectrum of the residual coloration for the same CTC experiment.

Both plots of the cancellation effectiveness and of the residual coloration show a maximal performance of the system in a band around 1500Hz. Coherently, an analysis of the phase delay of the (reverse of) the cancellation signal with respect to the crosstalk as a function of frequency shows that the alignment is perfect at about 1500Hz. For frequencies above 1500Hz the phase delay tends to stabilize at 0.5 samples and this explain the loss of performance at these frequencies. Below 1500Hz, the coefficients used for the head related delays introduce ineffectiveness, because of the

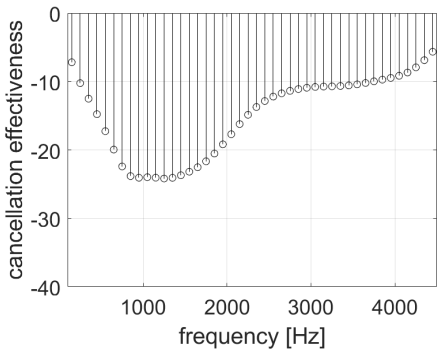

Fig. 3: Cancellation effectiveness (in dB) for the counterlateral truncation scheme at order 7.

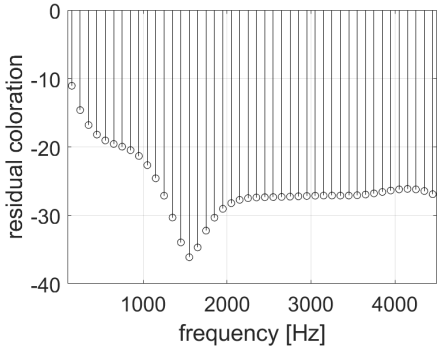

Fig. 4: Residual coloration (in dB) for the counterlateral truncation scheme at order 7.

diffractive nature of the propagation of the sound that becomes more and more important as the wavelength becomes larger with respect to the size of the head. A multi-band approach, which is an additional possible extension to the solution proposed in this paper, is a possible countermeasure to these performance losses, but is not mandatory in terms of user-perceived CTC effect.

Figure 5 and 6 provide the resulting cancellation effectiveness measured on one of the dummy head ears, for a series of 20 experiments performed by modifying the cancellation order N from 1 to 10, for both the ipsilateral and the counterlateral truncation schemes. For each cancellation order, the mean cancellation effectiveness in the frequency band between 750 Hz and

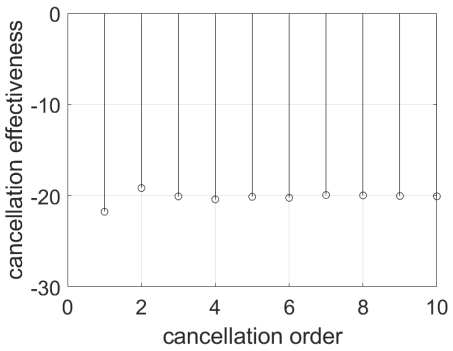

Fig. 5: Progression of cancellation effectiveness (in dB) for the ipsilateral truncation scheme.

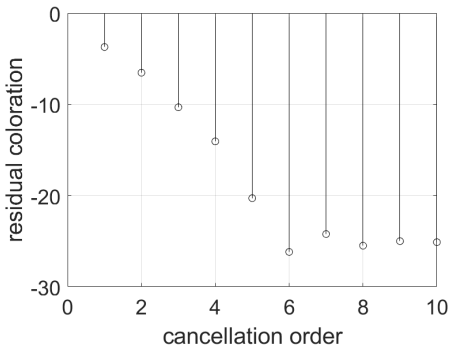

Fig. 7: Progression of coloration (in dB) for the ipsilateral truncation scheme.

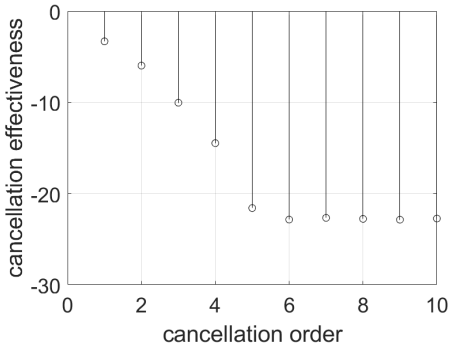

Fig. 6: Progression of cancellation effectiveness (in dB) for the counterlateral truncation scheme.

Fig. 8: Progression of coloration (in dB) for the counterlateral truncation scheme.

2000 Hz has been represented.

In figure 5 and 6, an important difference between the ipsilateral and counterlateral truncation schemes can be noticed. In the ipsilateral scheme, the cancellation effectiveness remains relatively stable from order 1 to order 10, whereas in the counterlateral scheme the cancellation effectiveness improves for the initial order increments, then finds its stability starting from cancellation order 6. The main reason of this progression, is the missing cancellation signal, consequence of the truncation, that generates a flawed CTC condition. In the experimented configuration, this side effect disappears relatively fast when incrementing the cancellation order and the CTC already finds it's stability at cancellation order 6.

Figure 7 and Figure 8 provide information about the intensity of the residual coloration measured at the target ear. It can be seen, that the residual coloration side-effect behaves in contraposition to the described flawed CTC condition visible at small orders. In the counterlateral truncation scheme, there's a stable amount of coloration for all cancellation orders, whereas in the ipsilateral scheme, the amount of coloration is high for small orders of CTC, but find its stability at cancellation order 6.

To provide additional performance validation, the experimental setup has then been reconfigured by shifting the position of the dummy head by 20 cm to the right and rotating its point view by 10 degrees, also in the right direction. This configuration produces larger in-

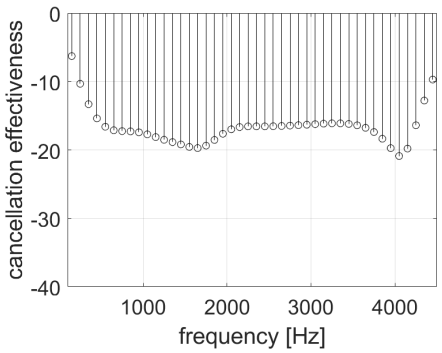

Fig. 9: Cancellation effectiveness (in dB) on the left ear of the shifted and rotated head, for the counterlateral truncation scheme at order 7.

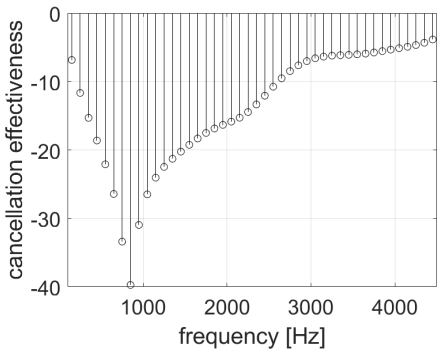

Fig. 10: Cancellation effectiveness (in dB) on the right ear of the shifted and rotated head, for the counterlateral truncation scheme at order 7.

cidence angles of the sound than the central position discussed above. Figure 9 and Figure 10 show the spectrum of the cancellation effectiveness measured on both ears, for the CTC experiment with order 7 and counterlateral truncation scheme. One can remark that the cancellation effectiveness, although not identical to the central symmetrical experimental configuration case, remains sufficiently strong in the frequency band of interest.

7 Summary and conclusions

In this paper, we have presented a novel time-domain approach to CTC, based on the modeling of the geometrical acoustics of the system. Our solution introduces the concept of Cancellation Complexes and their truncation and propose an elegant solution to the robustness issues of CTC systems. Laboratory experiments have confirmed the effectiveness of the proposed solution and its capacity to handle listeners in arbitrary spatial positions.

The promising results obtained so far open the door for multiple future research directions. Integration of dynamic user tracking and support for multiple users are already active research fields in our laboratory. Enhancement of the CTC effect by introduction of a multibands variant of our approach is also part of our ongoing research program.

This research was conducted with the support of 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] M., I. and T., M., “Stereophonic sound reproduction system,” 1978, uS Patent 4,118,599.

[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] Kirkeby, O. F., Nelson, P. A., and Hamada, H., “The stereo dipole : A virtual source imaging

system using two closely spaced loudspeakers," *Journal of The Audio Engineering Society*, 46, pp. 387–395, 1998.

[10] Park, C.-S., "Guaranteed-Stable Sliding DFT Algorithm With Minimal Computational Requirements," *IEEE Transactions on Signal Processing*, 65(20), pp. 5281–5288, 2017, doi:10.1109/TSP.2017.2726988.

[11] 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 China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pp. 212–216, 2014, doi:10.1109/ChinaSIP.2014.6889234.

[12] Bai, M. R. and Lee, C.-C., "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), pp. 1976–1989, 2006, doi:10.1121/1.2257986.

[13] Kirkeby, O., Nelson, P. A., and Hamada, H., "Local sound field reproduction using two closely spaced loudspeakers," *The Journal of the Acoustical Society of America*, 104(4), pp. 1973–1981, 1998, doi:10.1121/1.423763.

[14] Yang, J., Gan, W.-S., and Tan, S.-E., "Improved sound separation using three loudspeakers," *Acoustics Research Letters Online*, 4(2), pp. 47–52, 2003.

[15] Nawfal, I., Atkins, J., and Nimick, S., "Perceptual Evaluation of Loudspeaker Binaural Rendering Using a Linear Array," *Journal of The Audio Engineering Society*, 2014.

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

[17] Kim, L.-H., Lim, J.-S., and Sung, K.-M., "A new robust acoustic crosstalk cancellation method with sum and difference filter for 3D audio system," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, 85(9), pp. 2159–2163, 2002.

[18] Lentz, T. and Schmitz, O., "Realisation of an Adaptive Cross-talk Cancellation System for a Moving Listener," in *Audio Engineering Society Conference: 21st International Conference: Architectural Acoustics and Sound Reinforcement*, 2002.

[19] 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.

[20] Brown, C. P. and Duda, R. O., "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, IEEE, 1997.