Audio Engineering Society

Convention Paper

Presented at the 126th Convention
2009 May 7–10 Munich, Germany

Binaural reverberation using a modified Jot reverberator with frequency-dependent interaural coherence matching

Fritz Menzer[1], and Christof Faller[1]

[1] *Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

Correspondence should be addressed to Fritz Menzer (`fritz.menzer@epfl.ch`)

ABSTRACT

An extension of the Jot reverberator is presented, producing binaural late reverberation where the interaural coherence can be controlled as a function of frequency such that it matches the frequency-dependent interaural coherence of a reference binaural room impulse response (BRIR). The control of the interaural coherence is implemented using linear filters outside the reverberator's recursive loop. In the absence of a reference BRIR, these filters can be calculated from an HRTF set.

1. INTRODUCTION

Currently, many real-time 3D audio applications use head related transfer functions (HRTFs), despite the fact that binaural room impulse responses (BRIRs) would lead to a higher quality 3D audio rendering than HRTFs alone. BRIRs are not widely used because of the inconvenience of BRIR measurement and because of the high computational complexity of applying BRIRs to a sound signal in real-time.

Binaural reverberators have already been proposed in order to simulate BRIRs and to improve the quality of real-time 3D audio rendering. However, many of these approaches are limited to applying HRTFs to early reflections [3] and lack a specific processing for binaural late reverb (e.g. independent reverb is used for the left and right channels as opposed to reverb with the same interaural cues as the BRIR tail). More elaborate techniques [4] apply an average interaural coherence at all frequencies, but do not reproduce correctly the interaural coherence as a function of frequency.

It was already suggested in [5] that the frequency-dependent interaural coherence plays an important role for the spatial aspects of auditory perception.

Therefore, creating a reverberator that correctly reproduces this aspect of late reverberation was a promising option to improve the quality of binaural reverberation.

Jot [1,2] proposed a method to design reverberators based on feedback-delay networks, allowing to control the spectrum of the impulse response as well as the reverberation time as a function of frequency. In this paper we are proposing a binaural reverberator which is an extension of the Jot reverberator as described in [2], allowing also to match the interaural coherence as a function of frequency to the coherence of the true reverb contained in a reference BRIR. The proposed binaural reverberator plus HRTFs for the direct sound are a solution to simulate BRIRs with low computational complexity.

The paper is organized as follows: Section 2 describes the basic idea on how to use reverberators to model BRIRs. Estimating the needed parameters from reference BRIRs is described in Section 3. The design of a modified Jot reverberator suitable for BRIR modeling is described in Section 4. Experimental results are presented in Section 5. Section 6 explains how a binaural reverberator can be constructed from only a mono room impulse response and HRTFs. The conclusions are in Section 7.

2. MODELING BINAURAL ROOM IMPULSE RESPONSES USING REVERBERATORS

To enable separate processing of the direct sound and the reverberation tail of BRIRs, we decompose BRIRs into a direct sound and a tail part,

$$
\begin{aligned}
b_{\mathrm{L}}(n) &= b_{\mathrm{L,direct}} + b_{\mathrm{L,tail}} \\
b_{\mathrm{L}}(n) &= b_{\mathrm{R,direct}} + b_{\mathrm{R,tail}}\,,
\end{aligned}
\tag{1}
$$

where n is the discrete time index, the direct sound parts are denoted $b_{\mathrm{L,direct}}$ and $b_{\mathrm{R,direct}}$, and the remaining parts (tails) are denoted $b_{\mathrm{L,tail}}$ and $b_{\mathrm{R,tail}}$. Note that $b_{\mathrm{L,direct}}$ and $b_{\mathrm{R,direct}}$ are equivalent to HRTFs if the first reflection arrives sufficiently late (approximately 3 ms or more after the direct sound).

Figure 1 shows the decomposition of an example BRIR into direct and tail parts.

The proposed technique uses (1) to model BRIRs, where the early parts of the BRIR, $b_{\mathrm{L,direct}}$ and $b_{\mathrm{R,direct}}$, are implemented as FIR filters and the late parts of the BRIR, $b_{\mathrm{L,tail}}$ and $b_{\mathrm{R,tail}}$, are generated using a specially designed reverberator to avoid the computational complexity which would arise by directly convolving sound signals with the entire BRIR, which can be several seconds long.

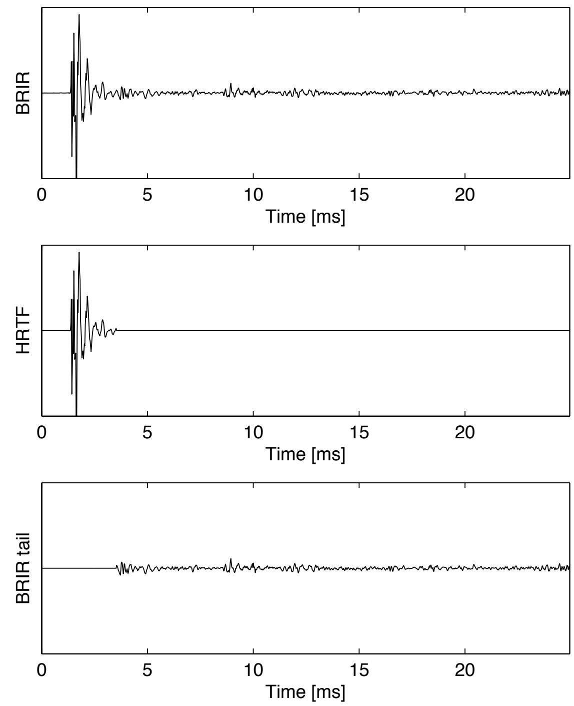

Fig. 1: Separation of BRIR into early and late parts. Note that for simplicity only the left channel of the BRIR is shown. **Top:** left channel of example BRIR. **Middle:** early part of the BRIR (corresponding to an HRTF in case only the direct sound is contained in it). **Bottom:** late part (tail) of the BRIR. All three plots have the same amplitude scaling.

The BRIR tails are modeled as

$$
\begin{aligned}
\hat{b}_{\mathrm{L,tail}}(n) &= h_L \star (u \star r_1 + v \star r_2)(n) \\
\hat{b}_{\mathrm{R,tail}}(n) &= h_R \star (u \star r_1 - v \star r_2)(n)\,,
\end{aligned}
\tag{2}
$$

where $\star$ denotes linear convolution, $r_1(n)$ and $r_2(n)$ are the uncorrelated impulse responses having the desired frequency-dependent reverberation time of the BRIR, $u(n)$ and $v(n)$ are used to perform the coherence matching, and $h_L(n)$ and $h_R(n)$ adjust the spectrum of the output signal to match the spectrum of the left and right channels of the reference BRIR.

In frequency domain, $h_L(n)$ and $h_R(n)$ are

$$\begin{aligned} H_L(\omega) &= \sqrt{\frac{P_L(\omega)}{T_r(\omega)}} \\ H_R(\omega) &= \sqrt{\frac{P_R(\omega)}{T_r(\omega)}}, \end{aligned} \tag{3}$$

where $T_r(\omega)$ is the frequency-dependent reverberation time as defined in [2]. Note also that $h_R(n)$ and $h_R(n)$ correspond to the tone correction filter $t(z)$ defined in [2]. In the following is described how to determine the filters $u(n)$ and $v(n)$. Note that these filters need to be short in order to achieve low computational complexity. In an FIR implementation, $u(n)$ and $v(n)$ may be truncated to suit eventual complexity requirements.

The desired filters in the frequency domain are written as

$$\begin{aligned} U(\omega) &= \sqrt{\frac{1+\Phi(\omega)}{2}} \\ V(\omega) &= \sqrt{\frac{1-\Phi(\omega)}{2}}, \end{aligned} \tag{4}$$

where $\Phi(\omega)$ is the coherence of the reference BRIR tail as a function of frequency.

Supposing that $r_1(n)$ and $r_2(n)$ are un-correlated, it can be shown from (2) and (4) that the coherence between $\hat{B}_{\mathrm{L,tail}}(\omega)$ and $\hat{B}_{\mathrm{R,tail}}(\omega)$ is $\Phi(\omega)$.

In order to be able to implement the late BRIR as defined in (2) efficiently, an artificial reverberator generating two uncorrelated reverb signals having the same frequency-dependent reverberation time is needed. The design of such an artificial reverberator yielding the desired reverberation signals $r_1(n)$ and $r_2(n)$ is described in Section 4.

3. ESTIMATING PARAMETERS FROM THE REFERENCE BRIRS

To obtain the filters $h_L(n)$, $h_R(n)$, $u(n)$ and $v(n)$, estimates of the left and right BRIR tail power spectra ($P_L(\omega)$ and $P_R(\omega)$) and of the coherence $\Phi(\omega)$ between the left and right BRIR tails are needed. In the following it is explained how to estimate these parameters, given the reference BRIR tails.

A short-time Fourier transform (STFT) is applied to overlapping blocks of the left and right BRIR tails, yielding $B_L(i,k)$ and $B_R(i,k)$, where i and k are the frequency and time indices, respectively. The reasons why we use an STFT are:

- The time-domain filters $u(n)$ and $v(n)$ need to be short in order to achieve low computational complexity. Thus, the reduced frequency resolution of the STFT compared to a single Fourier transform applied to the tail is enough.
- For the estimation of the coherence more than one sample per frequency is needed.

The power spectra and coherence are estimated as

$$\begin{aligned} P_L(i) &= \frac{1}{K}\sum_{k=1}^{K}|B_L(i,k)|^2 \\ P_R(i) &= \frac{1}{K}\sum_{k=1}^{K}|B_R(i,k)|^2 \\ \Phi(i) &= \frac{|\sum_{k=1}^{K} B_L(i,k)B_R(i,k)^\star|}{\sqrt{\sum_{k=1}^{K}|B_L(i,k)|^2\sum_{k=1}^{K}|B_R(i,k)|^2}}, \end{aligned} \tag{5}$$

where $|.|$ is the magnitude of a complex number, * denotes complex conjugate, and K is the number of STFT frames. If needed, the frequency resolution of $P_L(\omega)$ and $P_R(\omega)$ can be modified by interpolating the STFT bins to match the resolution of $T_r(\omega)$ in (3).

The top panel in Figure 2 shows the left and right power spectra of an example BRIR and the bottom panel shows the interaural coherence of the same BRIR as a function of frequency. The estimated $P_L(\omega)$, $P_R(\omega)$, $\Phi(\omega)$ can be frequency smoothed as much as needed in order to obtain short filters $u(n)$ and $v(n)$.

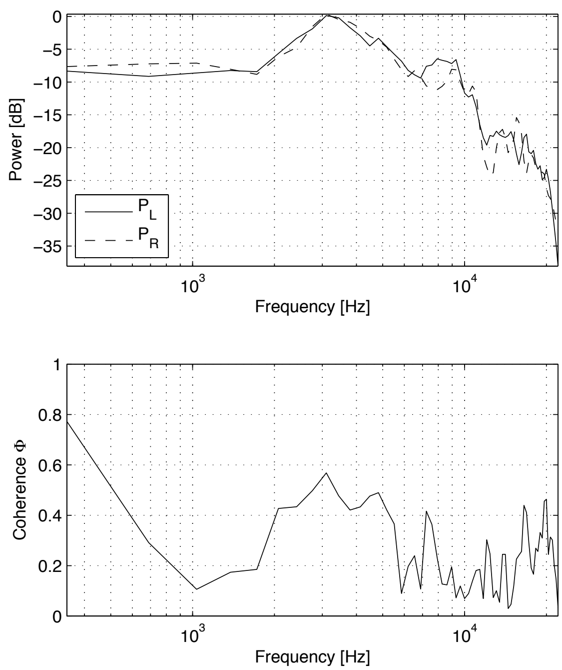

Fig. 2: Top: power spectrum estimated from example left and right BRIR tails. **Bottom:** corresponding estimated frequency-dependent interaural coherence.

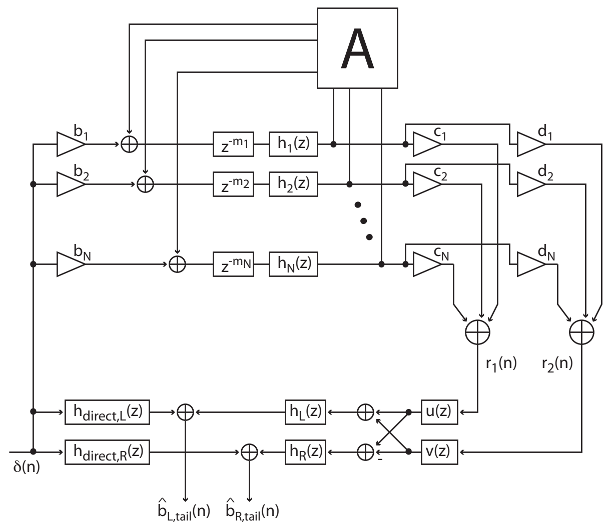

Fig. 3: Complete block diagram of binaural modified Jot reverberator. A denotes the mixing matrix as defined in [2] and $h_i(z)$ are the filters needed to adjust the frequency-dependent reverberation time. Names of signals were chosen for the case where the input is a single Dirac impulse and the output is the impulse response of the reverberator.

4. DESIGN OF AN UNCORRELATED TWO-CHANNEL REVERBERATOR

The complete reverberator as shown in Figure 3 is an extension of the Jot reverberator as described in [2], which was modified in order to produce a second, uncorrelated reverberation channel and to which the filters described in (2) were added.

The output signal in the original Jot reverberator is obtained by combining the N channels using a weights vector $\vec{c} = [c_1, c_2, \ldots, c_N]$, leading to an intermediate signal $r_1(n)$ and filtering this signal using a tone correction filter. The idea behind the structure of the reverberator proposed in this paper, as seen in Figure 3, is to use a second weights vector $\vec{d} = [d_1, d_2, \ldots, d_N]$ in order to combine the N channels of the reverberator to create an intermediate signal $r_2(n)$ which is not correlated with $r_1(n)$.

Under the hypothesis that the N channels of the reverberator produce uncorrelated Gaussian noise with equal amplitudes, it is obvious that the condition

$$\vec{c} \perp \vec{d} \tag{6}$$

is sufficient to assure that $r_1(n)$ and $r_2(n)$ are uncorrelated. In order to have the same energy in $r_1(n)$ and $r_2(n)$, it is necessary to impose also

$$||\vec{c}|| = ||\vec{d}||\,. \tag{7}$$

For $N = 6$, a practical example is

$$\begin{aligned} \vec{c} &= [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1] \\ \vec{d} &= [1 \quad \text{-1} \quad 1 \quad \text{-1} \quad 1 \quad \text{-1}]\ . \end{aligned} \tag{8}$$

In theory, trivial solutions to (6) and (7) like $\vec{c} = [100000]$ and $\vec{d} = [010000]$ are possible, but in practice these solutions are undesirable because they lead to a lower reflection density in the first part of the reverberation than the solutions given in (8).

The intermediate signals $r_1(n)$ and $r_2(n)$ are processed as defined in (2), where the filters $u(n)$, $v(n)$, $h_L(n)$ and $h_R(n)$ are implemented as FIR filters obtained by taking the inverse Fourier transforms of $U(\omega)$, $V(\omega)$, $H_L(\omega)$ and $H_R(\omega)$ as defined in (3) and (4). In practice, it is possible to truncate the obtained FIR filters, which makes the implementation more efficient, but in return leads to a loss of frequency resolution in the approximation of the left and right spectra as well as in the approximation of $\Phi(\omega)$.

To obtain the output signal of the binaural reverberator, the outputs of the filters $h_L(n)$ and $h_R(n)$ are added to the input signal convolved with the left and right channels of the HRTF corresponding to the direction of the direct sound.

5. EXPERIMENTAL VALIDATION

We applied our method to a BRIR measured in a lecture hall in our university and designed a binaural reverberator with $N = 6$. The frequency dependent interaural coherence of the reverberator was calculated using (5) and compared to the interaural coherence of the reference BRIR. As can be seen in Figure 4, the interaural coherence of the reverberator matches well the interaural coherence of the reference BRIR. However, the reverberator's interaural coherence is in general slightly higher. This effect proved to be systematic also for reverberators which were designed based on different BRIRs. A possible explanation is that the signals $r_1(n)$ and $r_2(n)$ are not completely uncorrelated, therefore leading to a higher correlation at the output. In this case it would be possible to compensate the correlation between $r_1(t)$ and $r_2(t)$ by introducing correction terms into (4).

Even though no extensive psychoacoustic tests have been performed, informal listening has lead us to the conclusion that a binaural reverberator with frequency dependent coherence matching performs better with respect to the goal of creating a realistic spatial image than a binaural reverberator with uncorrelated $\hat{b}_{\mathrm{L,tail}}(n)$ and $\hat{b}_{\mathrm{R,tail}}(n)$, and also better than a binaural reverberator with an average, frequency-independent coherence between $\hat{b}_{\mathrm{L,tail}}(n)$ and $\hat{b}_{\mathrm{R,tail}}(n)$.

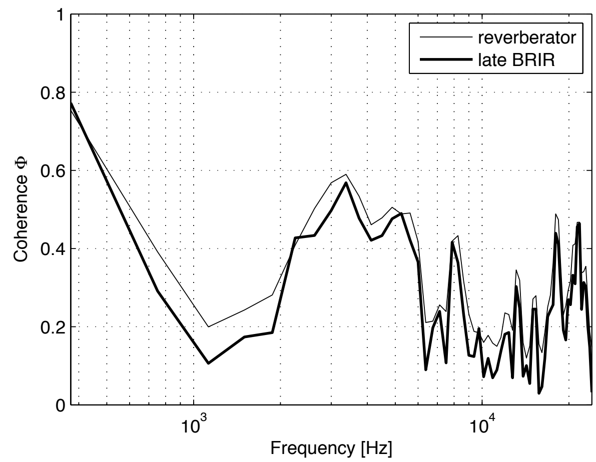

Fig. 4: Coherence of reference BRIR tail and of artificial reverberator. The coherence of the reverberator's impulse response follows the coherence of the reference late BRIR closely. However, a systematic bias towards higher coherence in the reverberator's impulse response can be observed. Unwanted correlation between $r_1(n)$ and $r_2(n)$ may be the cause of this increase in coherence.

6. BINAURAL REVERBERATION WITHOUT REFERENCE BRIRS

In the case where no reference BRIR is available, but a reference mono room impulse response $w(n)$ recorded with an omni microphone and an HRTF set are available, it is possible to design a binaural reverberator by estimating $P_L(\omega)$, $P_R(\omega)$ and $\Phi(\omega)$ from the HRTF set under the assumption that the reverberation tail consists of perfectly diffuse sound, using a method similar to the one proposed in [6].

In the following we assume the existence of an HRTF set whose individual left and right HRTFs are denoted $L_i(\omega)$ and $R_i(\omega)$, where $i \in \{1, 2, \cdots, I\}$ is the direction index and I is the number of HRTFs in the set.

By averaging over the individual HRTFs – rather than over the frames of an STFT – we obtain the desired quantities as follows:

$$
\begin{aligned}
P_L(\omega) &= \frac{|W_{\text{tail}}(\omega)|^2}{I} \sum_{i=1}^{I} |L_i(\omega)|^2 \\
P_R(\omega) &= \frac{|W_{\text{tail}}(\omega)|^2}{I} \sum_{i=1}^{I} |R_i(\omega)|^2 \\
\Phi(\omega) &= \frac{|\sum_{i=1}^{I} L_i(\omega) R_i^\star(\omega)|}{\sqrt{\sum_{i=1}^{I} |L_i(\omega)|^2 \sum_{i=1}^{I} |R_i(\omega)|^2}} ,
\end{aligned}
$$

where $W_{\text{tail}}(\omega)$ is the Fourier transform of the tail part of $w(n)$, obtained analogously to the BRIR tail parts in section 2.

7. CONCLUSIONS

A binaural reverberator based on the Jot reverberator was proposed, which allows to reproduce the frequency-dependent interaural coherence of a reference BRIR, in addition to the reference BRIR's frequency-dependent reverberation time and the left and right power spectra. This reverberator was implemented and its interaural coherence was found to match closely the interaural coherence of the reference BRIR.

Furthermore, informal listening suggests that for the impression of spaciousness, reproducing correctly the frequency-dependent interaural coherence is a significant improvement over reproducing only an average interaural coherence which does not depend on the frequency.

For the case where no reference BRIR is available, a method was presented to parametrize the binaural reverberator given only a mono room impulse response and an HRTF set.

ACKNOWLEDGEMENTS

We would like to thank Hervé Lissek and other people from EPFL's Electromagnetics and Acoustics Laboratory (LEMA) for their help with the BRIR measurements.

8. REFERENCES

[1] Jean-Marc Jot and Antoine Caigne, "Digital delay networks for designing artificial reverberators," in *Proc. 90th AES Convention*, 1991.

[2] Jean-Marc Jot, "An analysis/synthesis approach to real-time artificial reverberation," in *Proc. ICASSP-92*, 1992, vol. 2, pp. 221–224.

[3] N. Toma, M. Topa, and E. Szopos, "Aspects of reverberation algorithms," *ISSCS 2005.*, vol. 2, pp. 577–580 Vol. 2, July 2005.

[4] Jean-Marc Jot, Veronique Larcher, and Olivier Warusfel, "Digital signal processing issues in the context of binaural and transaural stereophony," in *Preprint 98th Conv. Aud. Eng. Soc.*, Feb. 1995.

[5] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Am.*, vol. 116, no. 5, pp. 3075–3089, Nov. 2004.

[6] F. Menzer and C. Faller, "Obtaining binaural room impulse responses from b-format impulse responses," in *Preprint 125th Conv. Aud. Eng. Soc.*, Oct. 2008.