HIGH PRECISION ROBUST MODELING OF LONG ROOM RESPONSES USING
WAVELET TRANSFORM

Sahar Hashemgeloogerdi and Mark Bocko*
Department of Electrical and Computer Engineering
University of Rochester

ABSTRACT
Modeling of a room impulse response (RIR) is required in
many audio processing applications; however, this is chal-
lenging since room responses are usually long and complex
in practice and drastically vary as the source and microphone
locations change. In this paper, a subband multichannel
modeling method is proposed, which is computationally effi-
cient, precise, and robust against RIR variations. A dual-tree
complex wavelet packet transform is utilized to decompose
a multichannel RIR into aliasing-free subband signals, and
low order adaptive Kautz filters are designed to model sub-
band signals using the poles common to the RIR channels. A
least-squares algorithm is introduced to efficiently estimate
the common poles at each subband. Experimental results in-
dicate that the proposed method accurately models long room
responses, while exhibiting significant robustness against
room response variations caused by changing the source and
microphone locations.

Index Terms— Wavelet transform, Kautz filter, least-
squares approximation, room acoustics.

1. INTRODUCTION
A room impulse response (RIR) describes the sound propaga-
tion characteristics between a source and a microphone placed
inside a room. Accurate modeling of an RIR is essential in
many acoustic signal processing applications such as room
acoustic virtualization [1] and acoustic feedback cancellation
[2]. In practice, modeling of an RIR is challenging since (i) an
RIR can be tens of thousands taps long, requiring high order
filters for accurate representation, and (ii) an RIR drastically
changes with slight variations in the source and microphone
locations.

Several methods for modeling of an RIR have been pro-
poused. All-zero modeling [3, 4] is the simplest method. For
a room with a relatively long reverberation time, however, an
all-zero model requires a large number of parameters. Since
poles represent room resonances with fewer parameters than
zeros, a pole-zero method [4] provides a more compact model
of an RIR. However, pole-zero models may require nonlinear
optimization, suffering from convergence to a local minima
[5]. An alternative to conventional modeling methods is the
Kautz filtering [6] which utilizes orthonormal basis functions
to provide a more precise model with fewer parameters.

In practice, due to the complex time-frequency structure
of a room response, a conventional modeling method or a
kautz filter may not accurately model the full audio frequency
range and temporal decay of a room response [6-9]. To al-
leviate this restriction, different subband modeling methods
have been proposed. A multirate system [7], the frequency
zooming ARMA model [8], and polyphase Kautz model [9]
have been shown to provide gentler performance as compared
to the fullband counterparts. These methods, however, do not
address the robustness against RIR variations. Models less
sensitive to RIR variations have been proposed in [10-12],
where the common acoustical poles of the room are utilized
to develop a pole-zero model [10, 11] or a model based on
orthonormal basis functions [12]. These models, however, do
not represent the room response over the full audio frequency
range and are limited to low frequency components of an RIR.

In this paper, we introduce a method for modeling of long
room responses. The proposed model exhibits high precision
and significant robustness against the RIR variations for the
full audio frequency range. Given a multichannel RIR, we
decompose the RIR into subband equivalent signals and de-
sign a low order adaptive Kautz filter at each subband before a
fullband signal is reconstructed. We utilize the dual-tree com-
plex wavelet packet transform [13, 14] to produce aliasing-
free subbands. We introduce a least-squares (LS) algorithm
for the efficient approximation of the poles common to the
RIR channels at each subband; common poles (CPs) enhance
the model robustness against the RIR variations. Estimated
poles are then fixed, and the model is validated with room
responses measured at source and microphone locations not
used for the approximation of the CPs. Experimental results
indicate that the proposed model provides a high precision
robust representation of long room responses by benefitting
from time-frequency decomposition and efficient approxima-
tion of CPs. In Section 2, we propose the subband multi-
channel RIR model. In Section 3, we validate the model with
experimental data. Conclusions are drawn in Section 4.
2. SUBBAND MULTICHANNEL RIR MODELING

A room response in the time domain can be divided into direct sound, early reflections, and late reverberations [15]. Alternatively, in the frequency domain, a room response can be characterized by discrete low frequency modes and diffuse overlapping modes [15, 16]. This complex time-frequency structure motivates the development of a subband multichannel RIR model. Since the wavelet oscillate locally, it requires an order of magnitude fewer coefficients than the Fourier basis to approximate within the same error [13]. We use the packet form of the discrete-time complex wavelet transform (DT-CWT) [14] to produce aliasing-free subbands and perfect reconstruction. Aliasing-free subbands are essential since aliasing causes erroneous results at the band edges [13].

2.1. Wavelet Decomposition

The DT-CWT consists of two wavelet transformers operating in parallel on a given signal [13]. We denote the wavelet associated with the first (second) wavelet filter bank (FB) as \( \psi(t) \) (\( \psi'(t) \)). Wavelet \( \psi(t) \) is defined as

\[
\psi(t) = \sqrt{2} \sum_n h_{lp}(n) \phi(2t - n),
\]

where \( \phi(t) = \sqrt{2} \sum_n h_{lp}(n) \phi(2t - n) \). Here, \( h_{lp}(n) \) and \( h_{hp}(n) \) represent a discrete-time low-pass and a discrete-time high-pass filter, respectively. Wavelet \( \psi'(t) \) is defined similarly in terms of \( \{h'_{lp}(n), h'_{hp}(n)\} \). For the ideal DT-CWT, wavelet \( \psi'(t) \) is the Hilbert transform of wavelet \( \psi(t) \),

\[
\psi'(t) = H\{\psi(t)\}.
\]

As shown in [13], [14], if \( h'_{lp}(n) \) is the half-sample delayed version of \( h_{lp}(n) \), wavelets produced by DT-CWT satisfy (2).

2.2. Subband Kautz Filters

Kautz filters are a special class of fixed-pole IIR filters, designed to produce orthonormal tap-output impulse responses [6]. The orthonormalization process provides control on individual resonances, enabling a Kautz filter to efficiently model an audio response. A Kautz filter is defined by a set of stable poles \( \mathbf{p} = \{p_m\}_{m=1}^N \) and a corresponding set of tap-output weights. A common assumption in modeling of a real room response is that the poles are real or complex conjugate [6]. For complex conjugate poles, a real Kautz filter formulation, shown in Fig. 1, prevents dealing with complex internal signals and filter weights. Normalization terms are given by [6]

\[
\alpha_m = \sqrt{(1 - \rho_m)(1 + \rho_m - \gamma_m)/2},
\]

\[
\beta_m = \sqrt{(1 - \rho_m)(1 + \rho_m + \gamma_m)/2},
\]

where \( \rho_m = |p_m|^2 \) and \( \gamma_m = -2Re\{p_m\} \). As illustrated in Fig. 1, the filter output is expressed as

\[
y(n, \mathbf{p}, \mathbf{w}) = \varphi^T(\mathbf{p}, n)\mathbf{w}(n),
\]

with \( \mathbf{w}(n) = [w_1(n), \ldots, w_{2N}(n)]^T \) and \( \varphi(\mathbf{p}, n) = [\varphi_1(n), \ldots, \varphi_{2N}(n)]^T \). Due to the short length of the subband signals, a low order Kautz filter at each subband is sufficient; furthermore, subbands are processed in parallel, enabling efficient computations. To enhance the robustness against room response variations, we use poles which are common to the RIRs measured at different source and microphone locations.

2.3. Least-Squares Approximation of common poles

Brandenstein and Unbehauen proposed an LS method, known as the BU method, for the FIR-to-IIR filter conversion [17], which produces unconditionally stable and optimal pole sets for a desired IIR filter order [17]. Inspired by the BU method, we propose an iterative LS algorithm that produces CPs at low expense. We refer to this algorithm as the CPBU method.

2.3.1. Problem formulation

The \( j \)-th subband equivalent signal of the \( i \)-th channel of a multichannel RTF, denoted as \( \mathbf{H}_{ij}(z) \) (\( i = 1, \ldots, M \) and \( j = 1, \ldots, J \)), is an FIR transfer function of length \( L \). The IIR approximation of \( \mathbf{H}_{ij}(z) \) in the least-squares sense is the IIR transfer function \( \mathbf{G}_{ij}(z) \approx \mathbf{H}_{ij}(z) \) of order \( N \) (\( N < L \)) such that

\[
E_{ij} = ||\Delta_{ij}(z)||_2 = ||\mathbf{H}_{ij}(z) - \mathbf{G}_{ij}(z)||_2
\]

is minimal, where \( || \cdot ||_2 \) denotes the \( l_2 \)-norm. To estimate CPs, which are independent of the source and microphone
locations, we define \( D_{1,j}(z) = \cdots = D_{M,j}(z) = D_j(z) \), where
\[
D_j(z) = \sum_{n=0}^{N} d_n z^{-n} = 1 + z^{-1} \sum_{n=0}^{N-1} d_{n+1} z^{-n} = 1 + z^{-1} \tilde{D}_j(z).
\]
Hence, approximating CPs in the least-squares sense means that we need to determine \( N \) real coefficients \( d_n \) (\( n = 1, \ldots, N \)) such that the \( l_2 \)-norm of the approximation error
\[
E_j = [E_{1,j}, \ldots, E_{M,j}]^T
\]
is minimized. As the CPs must be stable, the rational transfer function \( G_{ij}(z) \) is required to be analytic in \(|z| \geq 1\). Therefore, by applying the Walsh theorem [18], as discussed in [17], the difference function \( \Delta_{ij}(z) \) can be rewritten as
\[
\Delta_{ij}(z) = z^{-1} A_j(z) R_{ij}(z), \quad A_j(z) = \frac{z^{-N} \tilde{D}_j(z)}{D_j(z)}.
\]
where \( A_j(z) \) is an allpass filter, and \( R_{ij}(z) \) is an FIR transfer function with length \( L \) and real coefficients \( r_{ij,n} \) (\( n = 0, \ldots, L-1 \)) which are computed through allpass filtering of \( X_{ij}(z) = z^{-L} H_{ij}(z^{-1}) A_j(z) \) as
\[
U_{ij}(z) = \sum_{n=0}^{\infty} u_{ij}(n) z^{-n} = z^{-L} H_{ij}(z^{-1}).A_j(z), \quad (i = 1, \ldots, M)\]
\[
r_{ij,L-1-n} = u_{ij}(n) \quad (n = 0, \ldots, L-1).
\]
Using equations (6), (9), and (11), we have
\[
E_j^2 = \sum_{i=1}^{M} E_{ij}^2 = \sum_{i=1}^{M} ||\Delta_{ij}(z)||_2^2 = \sum_{i=1}^{M} \sum_{n=0}^{L-1} r_{ij,n}^2 = \sum_{i=1}^{M} \sum_{n=0}^{L-1} u_{ij}^2(n).
\]
Hence, least-squares approximation of the CPs is formulated as approximating \( D_j(z) \) such that the energy of \( u_j = [u_{1j}, \ldots, u_{Mj}]^T \) is minimal over the first \( L \) samples, where \( u_{ij} = [u_{ij}(0), \ldots, u_{ij}(L-1)] \) (\( i = 1, \ldots, M \)).

2.3.2. Approximation algorithm

We define the digital filter
\[
A_j^{(k)}(z) = \frac{z^{-N} D_j^{(k)}(z)}{D_j^{(k-1)}(z)}
\]
that approaches an allpass if \(|D_j^{(k)}(z) - D_j^{(k-1)}(z)||_2 \to 0 \) for \( k \to \infty \). Filtering \( X_{ij}(z) = z^{-L} H_{ij}(z^{-1}) A_j(z) \) leads to
\[
U_{ij}^{(k)}(z) = A_j^{(k)}(z) X_{ij}(z) = z^{-N} D_j^{(k)}(z) X_{ij}(z), \quad \text{(14)}
\]
where \( X_{ij}(z) = z^{-L} H_{ij}(z^{-1}) A_j(z) \). Substituting (7) in (14) for \( i = 1, \ldots, M \) leads to
\[
\begin{cases}
X_{1,j}^{(k)}(z) z^{-(N-1)} \tilde{D}_j^{(k)}(z) = U_{1,j}^{(k)}(z) - z^{-N} X_{1,j}(z) \\
\vdots \\
X_{M,j}^{(k)}(z) z^{-(N-1)} \tilde{D}_j^{(k)}(z) = U_{M,j}^{(k)}(z) - z^{-N} X_{M,j}(z).
\end{cases}
\]
Equating the coefficients of \( z^0, z^{-1}, \ldots, z^{-(L-1)} \) on both sides of the equations in (15), we have
\[
C_j^{(k)} \cdot d_j^{(k)} = u_j^{(k)} + b_j^{(k)}
\]
where
\[
C_j^{(k)} = [C_{1,j}^{(k)}, \ldots, C_{M,j}^{(k)}]^T, \quad d_j^{(k)} = [d_{1,j}^{(k)}, \ldots, d_{M,j}^{(k)}]^T, \\
b_j^{(k)} = [b_{1,j}^{(k)}, \ldots, b_{M,j}^{(k)}]^T, \quad u_j^{(k)} = [u_{1,j}^{(k)}, \ldots, u_{M,j}^{(k)}]^T, \\
b_j^{(k)} = [0, \ldots, 0, x_{ij}^{(0)}, \ldots, x_{ij}^{(L-1)}(L - N - 1)]
\]
\[
C_j^{(k)} = \begin{bmatrix}
\hat{x}_{ij}^{(0)}(1) & 0 & \cdots & 0 \\
\hat{x}_{ij}^{(1)}(0) & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
\hat{x}_{ij}^{(L-1)}(L-1) & \cdots & \cdots & \hat{x}_{ij}^{(L-N)}(L-N)
\end{bmatrix}.
\]
Solving \( C_j^{(k)} \cdot d_j^{(k)} = b_j^{(k)} \) in the LS sense, a vector \( d_j^{(k)} \) is obtained that minimizes the norm of \( u_j^{(k)} = C_j^{(k)} \cdot d_j^{(k)} - b_j^{(k)} \). Hence, we have found \( D_j^{(k)}(z) \) that minimizes (12).
We repeat the procedure as listed in Algorithm 1, creating a sequence of \( D_j^{(k)}(z) \) polynomials. Roots of the \( D_j^{(k)}(z) \) with the minimum error in sequence give the CPs.

**Algorithm I CPBU Algorithm**

**Input:** Subband signals \{\( H_{ij}(z) \)}_{i=1}^{M}, number of CPs \( N \)

1: \( D_j^{(0)} \leftarrow 1 \)
2: for \( k = 1, 2, \ldots, K \) do
3: \hspace{0.5cm} for all \( i \) do
4: \hspace{1cm} Update \( X_{ij}^{(k)}(z) = z^{-L} H_{ij}(z^{-1}) D_j^{(k-1)}(z) \)
5: \hspace{0.5cm} end for
6: Update \( C_j^{(k)} \) and \( b_j^{(k)} \) using \( X_{ij}^{(k)}(z) \) as in (17)
7: Solve \( C_j^{(k)} \cdot d_j^{(k)} = b_j^{(k)} \) in the least-squares sense
8: Store \( d_j^{(k)} \) and \( u_j^{(k)} \)
9: end for
10: Choose \( d_j \) corresponding to the \( u_j \) with the minimum norm
11: Create the \( D_j(z) \) polynomial utilizing elements of \( d_j \) as in (7)
12: Compute roots \( p_{1,j}, \ldots, p_{N,j} \) of polynomial \( D_j(z) \)

**Output:** The \( p_{1,j}, \ldots, p_{N,j} \) represent the CPs

2.4. Filter Weights Adaption

Once the CPs are found from the measured data, the filter weights \( w(n) \) can be adaptively computed by a standard algorithm such as Kalman filter, recursive least squares, and normalized least mean squares (NLMS) algorithm [19], [20]. For simplicity, we choose the NLMS with the adaptation rule
\[
\dot{w}(n+1) = \hat{w}(n) + \frac{\mu}{||\varphi(n, .)||^2} \varphi(n, .)(y(n) - \varphi^T(n, .) \hat{w}(n))
\]
where \( \mu \in (0, 2) \) is a gain. Linearity in filter weights leads to global convergence under the same conditions as for an FIR filter with the same number of parameters, and orthonormality assures faster convergence of the adaptation algorithm [20].
Fig. 2. Error of the SBCP-Kautz and CP-Kautz methods at different locations of the source and microphone.

3. EXPERIMENTAL RESULTS

Experimental results are presented to demonstrate the performance of the proposed subband multichannel Kautz method developed using the CPs (SBCP-Kautz). Experiments aim to evaluate the precision and robustness of the SBCP-Kautz method in both time and frequency domains in comparison to the fullband counterpart (CP-Kautz). To study the performance of the SBCP-Kautz method in modeling of the \( t \)-th channel of a multichannel RIR, the normalized mean square error (NMSE) is defined as

\[
\text{NMSE}_i = 20 \log 10 \frac{||h_i - \hat{h}_i||_2}{||h_i||_2} \quad (\text{dB}),
\]

where \( h_i \) and \( \hat{h}_i \) denote the measured and modeled RIRs.

Experiments are performed using the MARDY database [21]. The RIRs were measured from three sources (placed at left (L), center (C), and right (R)) to a linear array of microphones placed at three different locations in a room with reflective panels. The source-to-microphone distances varied between 1 m and 3 m (1 m increments). The \( T_{60} \) is 447 ms. Each RIR has a length of 65536 samples with sample rate \( f_s = 48 \) KHz. The SBCP-Kautz with 64 subbands is utilized, where each subband is modeled with 16 CP pairs (32-nth order filter). The CPBU algorithm (10 iterations) is used to estimate CPs from training data comprised of 7 RIRs measured from the source L to the microphone array placed at a 2 m from the source. The CPs are then fixed, and the SBCP-Kautz is applied to model room responses not used in the training step.

As illustrated in Fig. 2, the SBCP-Kautz method achieves nearly perfect modeling (NMSE \( \approx -80 \) dB), independent of the source and microphone locations. To provide a comparison, the CP-Kautz with a 64-th order filter was utilized to model the same room responses. The CP-Kautz is computationally expensive and gives much lower accuracy (NMSE \( \approx -11 \) to \(-17 \) dB) which, depending on the source and microphone locations, varies by more than 50%. Higher orders of a filter slightly enhance the accuracy of the CP-Kautz at the expense of highly increased computational load.

To evaluate the performance of the SBCP-Kautz method in the frequency domain, an arbitrary channel from the described database is considered. Fig. 3 shows the channel frequency response and corresponding modeled response, followed by the error signal. The SBCP-Kautz models the room response over the full audio frequency range with almost no degradation, while benefiting from the low order filters.

4. CONCLUSIONS

A method for modeling of long room responses has been introduced, where the RIR is decomposed into aliasing-free subband signals and a low order adaptive Kautz filter is designed to model each subband using CPs. A least-squares algorithm is introduced to efficiently estimate the CPs. Experimental results indicate that the proposed method (i) precisely models RIRs over the entire audio frequency range, and (ii) is robust against RIR variations. Computational efficiencies afforded by employing the subband method make it practical to implement the proposed method on modest signal processing platforms.
5. REFERENCES


