Friedrich-Alexander-Universität Erlangen-Nürnberg

**Lehrstuhl für Multimediakommunikation und Signalverarbeitung**

Prof. Dr.-Ing. Walter Kellermann

Research Internship

# Blind Estimation of the Subband Reverberation Time

Shrishti Saha Shetu

April 2020

Supervisor: Dr.-Ing. Heinrich Löllmann

# Declaration

I confirm that I have written this report unaided and without using sources other than those listed and that this thesis has never been submitted to another examination authority and accepted as part of an examination achievement, neither in this form nor in a similar form. All content that was taken from a third party either verbatim or in substance has been acknowledged as such.

| | |
|---|---|
| Place, Date | Shrishti Saha Shetu |

# Contents

# Acronym

| | |
|---|---|
| **RT** | Reverberation Time |
| **RIR** | Room Impulse Response |
| **DCT** | Discrete Cosine Transform |
| **ANN** | Artificial Neural Network |
| **DNN** | Deep Neural Network |
| **CNN** | Convolutional Neural Network |
| **ACE** | Acoustic Characterisation of Environments |
| **MAE** | Mean Absolute Error |
| **MSE** | Mean Square Error |
| **LSE** | Least Square Error |
| **ASR** | Automatic Speech Recognition |
| **DRR** | Direct-to-Reverberant Ratio |

# Abstract

Reverberation, in psycho-acoustics and acoustics, refers to the prolongation of sound in an enclosure after the sound source is switched off. A common measure of this property is the reverberation time (RT). Blind estimation of the subband RT is still a challenging task as the determination of the ground truth using the Schroeder method is already a demanding and tedious task.

The aim of this research work was to investigate known and novel model-based approaches for the estimation of the subband RT.

In a first step, a recently presented approach to estimate the subband RT by extrapolating the RTs for the higher subband from the estimates of the lower subbands [2] has been investigated and compared with the related approaches presented in [1]. It turned out that the model described in [1] achieves the lowest average error per subband for all the tested room impulse respones.

The development of a model for the subband RT requires large databases with ground truth data for the subband RTs. However, many databases with room impulse responses (RIRs) do not contain the ground truth data. Therefore, an approach to estimate the subband RT from a given RIR by means of the Schroeder method has been developed.

In a next step, two model-based approaches [3] using discrete cosine transform (DCT) filterbanks were evaluated. Based on this and previous results, a polynomial regression model for the subband RTs has been investigated. The analysis results show that, it's possible to approximate a model using DCT filterbanks, but it highly depends on the acoustic properties of the room and fullband RT range, which makes it quite tough to find an universal regression model with the known approaches. The mean absolute average error per subband for a 30 channel DCT filterbank considering all fullband RT range was 0.0400s for higher subands starting from the subband number 15 and 0.1010s considering all the subbands.

Finally, the use of artificial neural networks (ANNs) for the estimation of fullband RT from subband RTs as well as the estimation of the subband RT from a given RIR have been investigated. It turned out that the ANN model using a 30 channel DCT filterbank to estimate the fullband RT from subband RTs acheives 0.0048s mean square

error (MSE) and for an octave filterbank the MSE was 0.0062s. On the other hand, the mean square error per subband for the estimation of the subband RT from a given RIR was 0.0320s using a 30 channel DCT filterbank .

# Chapter 1

# Introduction

The reverberation time (RT) is an important room acoustical parameter. It is typically defined as the time interval in which the sound energy drops by 60dB after switching of the excitation source. In literature, it is also defined as T60. The RT can be used to identify the acoustic environment, and the knowledge about the RT can be exploited for enhanced automatic speech recognition (ASR), as well as speech dereverberation [1]. Several algorithms have been developed till date for fullband and subband RT estimations. Unfortunately, the estimation accuracy for the frequency dependent reverberation time is lower than the fullband reverberation time estimation as discussed in [1].

A common approach for blind estimation of the fullband RT is to use maximum likelihood (ML) estimation [2]. However it has been shown that the estimation accuracy for the frequency dependent RT using an octave filterbank by ML estimation is not so promising due to the smaller bandwidth of the subband filters and thus the lower signal energy, later discussed in Chapter 2.

Previously, to alleviate the problem that arises from ML estimation for the subband RT, a Rayleigh model-based approach was proposed [2]. In this approach, the subband estimates for the higher subbands are extrapolated from the more reliable estimates of the lower subbands by means of a model function for the subbands.

Recently a hybrid approach for the blind estimation of the subband RT has been proposed [1], where the RTs for the higher subbands from 4 to 20kHz are extrapolated from the blind RT subband estimates of the lower frequency bands. The assumption was that the RT decreases monotonically from 4 to 20 kHz due to the material and air absorption. This motivated to build a second order polynomial model for the RT estimation from mid- to high-frequencies. This model combines the ML estimation method for determining the subband RTs in the frequency range 1 to 4 kHz.

A comparison of these published approaches is also presented in this report. The analytical results show that the hybrid model-based approach has a lower bias and variance than the other approaches for most of the subbands, the per subband error

is also the lowest using this model for most cases. However these results give an indication that a proper model-based algorithm can help in estimating the frequency dependent RT.

A challenge in the development of learning-based approaches for the subband RT estimation is the creation of RIR databases with ground truth data for the subband RT. The Schroeder method is widely accepted as a way to find the ground truth RT from a RIR [6]. But there are some limitations of the Schroeder method itself, as the Schroeder method fits only well above the Schroeder frequency. Determining the RT from a given RIR manually using the Schroeder method is a tedious task, as for a proper model we will need a huge and varying dataset and the subsequent ground truth RT for all the subbands. But till date there are no methods have been published for automatic subband RT determination using the Schroeder method. This problem is also addressed in this research work and a new implementation for automatic RT calculation in subbands using the Schroeder method is proposed.

The use of a DCT filterbank may ease the effect of less energy in lower subbands, but it's needed to be investigated whether it is possible to approximate a universal model for all the subbands with current analytical approaches. So two model based approaches using DCT filterbanks for the subband RT estimation were also evaluated [3].

Therefore, a first order polynomial model using a 30 channel DCT filterbank has been developed. The outcome of this investigation also manifested that these kinds of analytical models are highly dependent on acoustical parameters. So instead of an analytical model, a learning-based approach has been investigated to determine the subband RTs.

In literature, it is quite frequently discussed that there is a certain tendency for the higher subband RTs. In [1] and [3], this effect was also investigated. But there is no known literature which examined, if it is possible to estimate the fullband RT from the subband estimations. This will also re-ensure the assumption of a certain distribution of RT in subbands. An analytical or learning-based model can investigate the relationship between subband and fullband RT estimations. A new approach based on an ANN to establish this relationship between subband and fullband RT estimations is also presented in this report. Finally, a deep neural network (DNN) model for the subband RT estimation directly from a given RIR is proposed.

In Chapter 2, common model-based approaches are discussed. A comparison among these models has also been investigated in this chapter. In Chapter 3, the Schroeder method is described briefly. A proposed automatic Schroeder method is also explained

here. In Chapter 4, two analytical model-based approaches for the subband RT estimations and a polynomial regression model for the subband RT estimation are discussed. An ANN approach for the fullband RT estimation from the subband estimations and the blind subband RT estimation directly from the RIRs are described as well.

# Chapter 2

# Common Model-Based Approaches

As discussed already in the introduction, there are different approaches for subband RT estimation. A short overview of some of these approaches is provided in this chapter.

## 2.1 Maximum Likelihood Method

In [4] the sound decay $d(k)$ within a speech pause caused by the room reverberation has been modeled using a discrete random process:

$$d(k) = A_r m(k) e^{-\rho k / f_s} u(k) \ . \tag{2.1}$$

Here $A_r$ denotes the amplitude of the sound decay and $u(k)$ represents the unit step function, $f_s$ is the sampling frequency and $m(k)$ represents a random sequence with zero mean and normal distribution. $\rho$ is the decay rate, which is related to the T60 by:

$$\rho = \frac{3}{\log_{10}(e) T_{60}} \ . \tag{2.2}$$

The probability density function for this decay rate can be expressed as following:

$$P_{d(k)}(x) = \frac{1}{\sqrt{2\pi}\epsilon(k)} e^{-\frac{x^2}{2\epsilon^2(k)}}; \ . \tag{2.3}$$

Then the decay rate $\rho$ can be obtained by finding the maximum of the log-likelihood function (natural logarithm of the likelihood function) from a given sequence $d(k)$ of length $N$ [4].

$$L(\rho) = \frac{-N}{2} \left[ (N-1) \ln(a) + \ln \left[ \frac{2\pi}{N} \sum_{i=0}^{N-1} a^{-2i} d^2(i) \right] + 1 \right] \ . \tag{2.4}$$

Then the estimation can be written as following:

$$\tilde{\rho}^{ML} = \max_{\rho} L(\rho); \ . \tag{2.5}$$

Now by using (2.2), the RT after some post-statistical processing for a given reverberant speech signal or RIR can be estimated [5].

## 2.2 Rayleigh Model-Based Method

The estimation of the subband RT with an octave filterbank is especially difficult for the upper subbands, even if the RIR is given [2]. This problem becomes even more pronounced w.r.t. a blind estimation, and calculating the subband RT by means of a common algorithm (such as, maximum likelihood estimation and the Schroeder method etc.) leads to a high estimation error especially for the upper subbands using an octave filterbank. To alleviate this problem, the following approach has been developed in [2]. The subband estimates for the upper subbands are extrapolated from the more reliable estimates of the lower subbands (for the subband index: $n_{\text{oct}}$, . . . ,$N_{\text{oct}}$) by means of a model function for the subband RT. For this purpose, the following simple model for the frequency dependent RT is devised. Inspection of the ground truth data for the subband RT $T_{60}(\mu)$ of the ACE (acoustic characterisation of environments) challenge DEV database [6] has shown that the frequency-dependency of the subband RT can be often roughly approximated by a function similar to that of a scaled Rayleigh distribution with an offset $m_0$ [2]. This leads to the following model function:

$$f_{mod}(\mu, b) = \frac{\mu}{\alpha b^2} \exp\left(\frac{-\mu^2}{2\alpha^2 b^2}\right) + m_0 \ . \tag{2.6}$$

$$m_0 = \frac{1}{N_{oct} - n_{oct} + 1} \sum_{\mu=n_{oct}}^{N_{oct}} \hat{T}_{60}^{oct}(\mu) \ . \tag{2.7}$$

Here $\mu$ is the subband index and $\alpha$= 7.50. The optimal scaling factor $b_{opt}$ is calculated by minimizing a least-square error between the model function and the RT estimates for the lower subbands. Then the subband RT estimates are finally given by:

$$\hat{T}_{60}^{(\text{model})}(\mu) = f_{mod}(\mu, b_{opt}) \quad \forall \mu \in (1, \ldots, N_{oct}) \tag{2.8}$$

## 2.3 A Hybrid Method

In [1], it was observed that the T60 decreases monotonically from 4 to 20 kHz due to the material and air absorption. The authors utilized this effect to build a model for the T60 from mid- to high-frequencies. In this study, a 2nd-order polynomial function is applied as a model to predict the T60 from mid- to high-frequencies, given as follows:

$$\hat{T}_{60}^{(\text{model})}(f_c) = a(f_c - 4\text{kHz})^2 + b(f_c - 4\text{kHz}) + T_{60,4\text{kHz}} \quad \text{for} \quad 4\text{kHz} \leq f_c \leq 20\text{kHz} \tag{2.9}$$

where $\hat{T}_{60}^{(\text{model})}(f_c)$ is the modeled reverberation time, $T_{60,4\text{kHz}}$ is the reverberation time blindly estimated at 4 kHz and $f_c$ is the center frequency of the used gammatone filterbank. The two model parameters $a$ and $b$ can be estimated by solving a nonlinear least squares optimization problem. In the case of an unknown room, these two parameters can be estimated from the more reliable estimates of the T60 in mid-frequency channels. For this purpose, the averaged reverberation time from 1 to 4 kHz, $T_{60,4\text{kHz}}$ is mapped to the model parameters $a$ and $b$ by a 2nd-order polynomial function. The mapping functions can be written as follows:

$$a = c_1 \bar{T}_{60,1-4\text{kHz}}^2 + d_1 \bar{T}_{60,1-4\text{kHz}} + e1. \tag{2.10}$$

$$b = c_2 \bar{T}_{60,1-4\text{kHz}}^2 + d_2 \bar{T}_{60,1-4\text{kHz}} + e2. \tag{2.11}$$

In [1], the parameters $c_1$, $c_2$, $d_1$, $d_2$, $e_1$ and $e_2$ were fitted to the estimated model parameters $a$ and $b$, and the $\bar{T}_{60,1-4\text{kHz}}$ for six different rooms by using the trust-region algorithm. In our evaluation, we used the Matlab function $polyfit$. First we stored all the parameters of (2.9) for six different RIRs from the ACE dev dataset in Matlab arrays. Then we calculated the parameters $a$ and $b$ using the $polyfit$ function. After that we took the mean of the estimated RT till subband number 23 (total 30 subbands). Then as before we stored the parameters of (2.10) and (2.11) in a Matlab array. Finally we again used the $polyfit$ function and the already calculated parameters $a$ and $b$ to get the model coefficients $c_1$, $c_2$, $d_1$, $d_2$, $e_1$ and $e_2$.

Fig. 2.1 shows the block diagram of the proposed algorithm for blind estimation of the T60 from 100 Hz to 20 kHz. In [1], a gammatone filterbank was used, where an octave filterbank was used in this work, and the T60 in each frequency channel is blindly estimated by the ML method . In the case of a noisy environment, the accuracy of the blindly estimated T60 might be reduced. Therefore, in [1], a median filter is

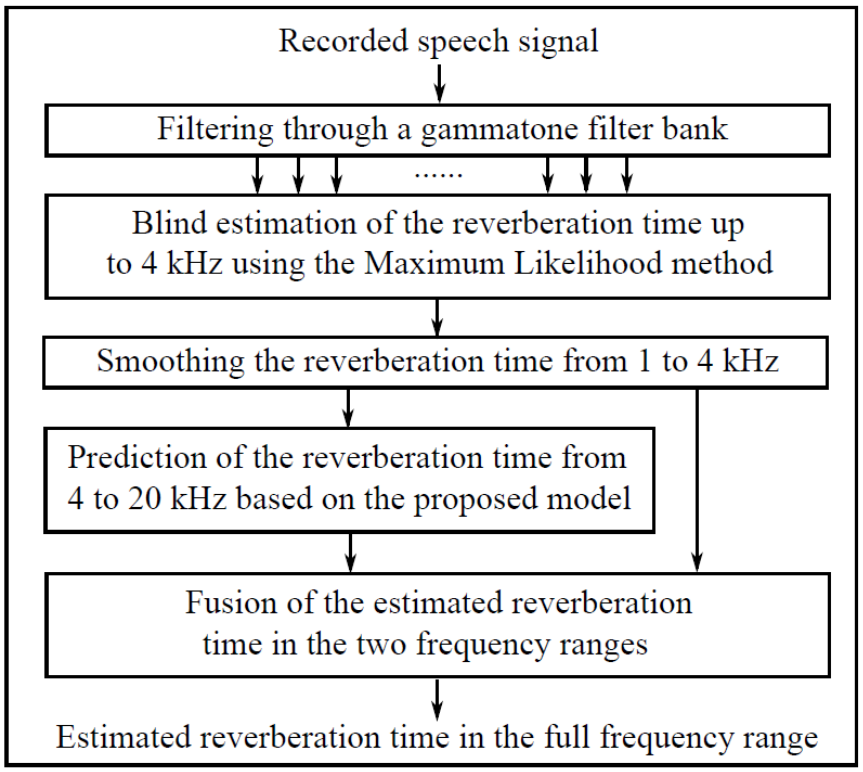


Figure 2.1: Structure of the method for blind estimation of the frequency dependent reverberation time, taken from [1].

applied as a smoothing filter for the estimated T60 from 1 to 4 kHz, where a third order median filter (provided by Matlab) was used in this work. Then the averaged reverberation time from 1 to 4 kHz, $T_{60,1-4\text{kHz}}$ was calculated to obtain the model parameters $a$ and $b$ according to (2.10) and (2.11). After that, the T60 from mid- to high frequencies (4-20 kHz) was predicted based on the estimated reverberation time at 4 kHz $\hat{T}_{60;4\text{kHz}}$ and the obtained model parameters $a$ and $b$ by using (2.9). Finally, the T60 from 100 Hz to 20 kHz can be determined.

## 2.4 Results

In Fig. 2.2, we can see a visualization for the T60 estimation in subbands with different models. This example shows a rare case as there is a sudden jump in T60 in the upper subbands. The Hybrid model cannot approximate well this phenomena as it assumes a monotonic decrease of the T60 for the upper subbands ranges from 4 kHz to 20 kHz. In contrast, the Rayleigh model doesn't suffer much from this effect, though it cannot

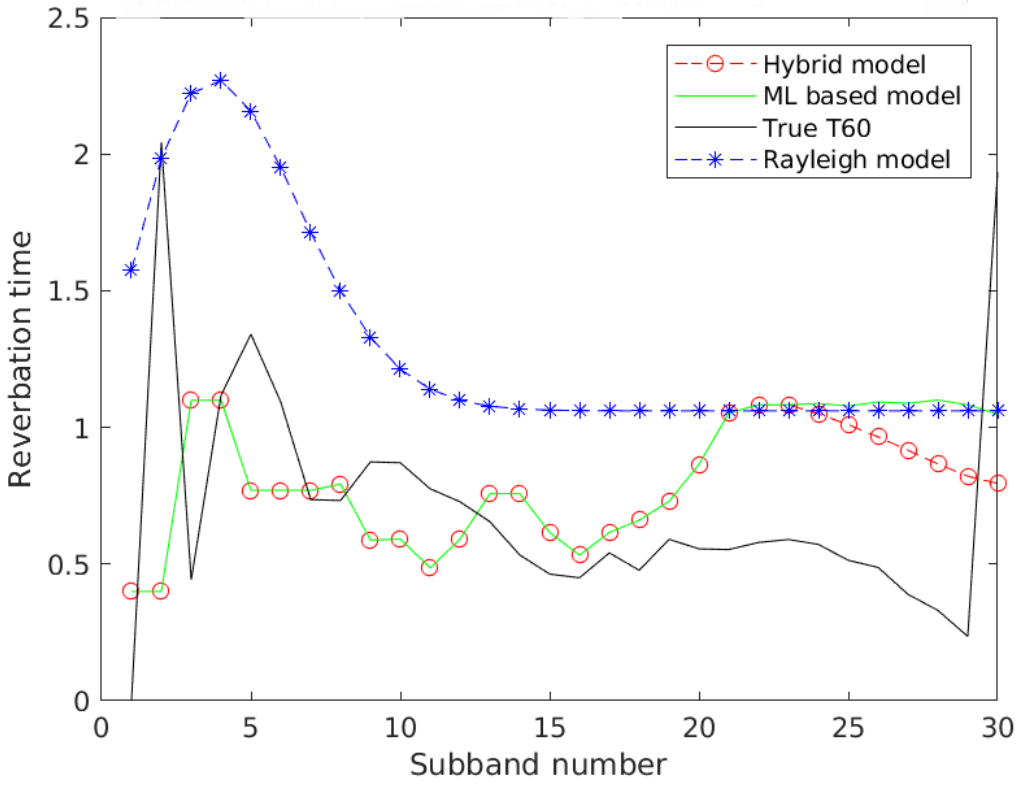


Figure 2.2: T60 estimation in subbands with different models using an octave filterbank.

sufficiently approximate the T60 in the lower subbands .

Furthermore, we calculated the average estimation error for the RIRs of the ACE challenge dataset with all the specified models using the following formula:

$$\mathrm{AE} = \frac{\sum_{i=1}^{N_{fc}} \mid \tilde{T}_{60}(f_{c,i}) - T_{60}(f_{c,i}) \mid}{N_{f_c}}. \tag{2.12}$$

Here AE is the average absolute error per subband, $\tilde{T}_{60}(f_{c,i})$ is the estimated RT in the specified frequency band, $T_{60}(f_{c,i})$ is the true RT in the specified frequency band and $N_{f_c}$ is the number of subbands, which was equal to 30 for these experiments.

The estimated RT in subbands using different models is plotted in Fig. 2.3. It shows that the Hybrid model achieves the lowest error for all the considered RIRs. Though the error margin between the ML and the Hybrid model is not so wide, as the Hybrid model itself uses the ML model for the estimation of T60 for the subbands ranges between 1 to 4kHz . The Rayleigh model shows the highest estimation error in this case, which suggests that for the used room impulses the model doesn't fit well.

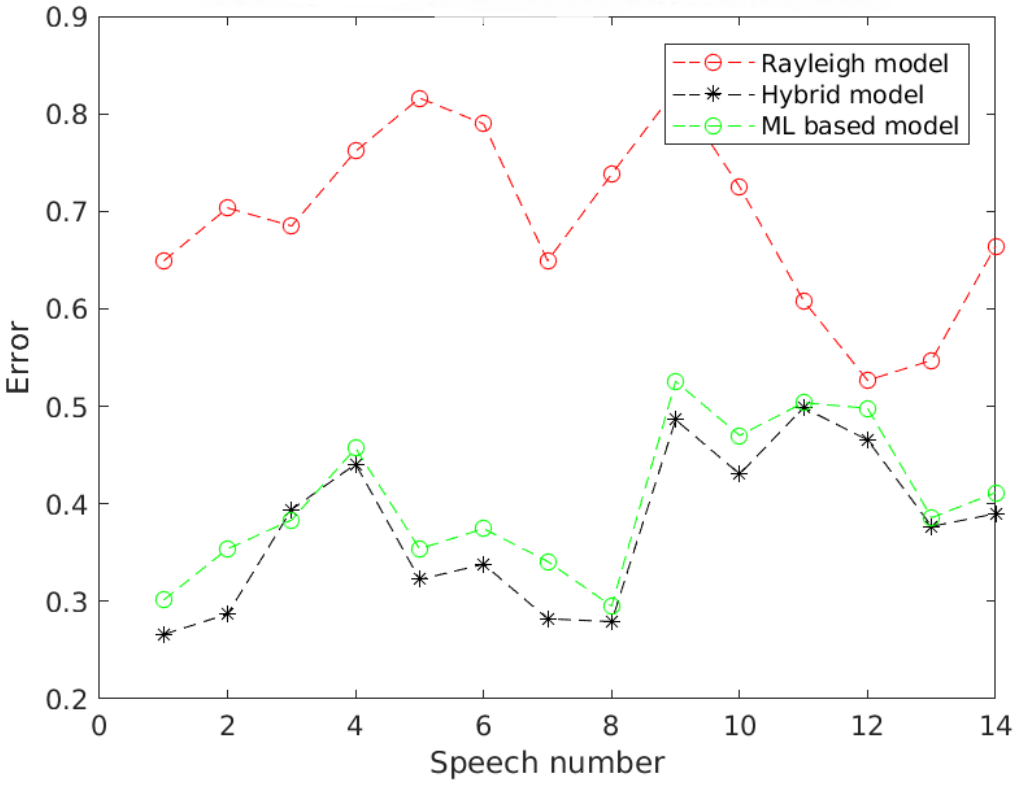


Figure 2.3: Error comparison for different models.

# Chapter 3

# The Schroeder Method

As already pointed out in the introduction, a problem in finding a model for the frequency dependent RT is a lack of labeled data. Many databases with RIR are published but the ground-truth RT in the subband domain is not always provided. The Schroeder method can be used but finding a suitable decay region is difficult. Therefore, the following approach has been developed.

The Schroeder method [7] has the constraint of the Schroeder frequency [8]. Since this method only fits well above the Schroeder frequency, it is very difficult to determine the T60 for the lower subbands with this method.. The Schroeder frequency is given by :

$$f_s = 2000\sqrt{\frac{T}{V}}. \tag{3.1}$$

Here V is the room volume ($m^3$), $T$ is the reverberation time (sec.) [9] and $f_s$ is the Schroeder frequency (Hz).

The reverberation time formula of Sabine [9] is:

$$RT = 0.161\frac{V}{\alpha A}. \tag{3.2}$$

Here A is the surface area ($m^2$), and $\alpha$ is the average absorption coefficient of the surfaces.

Usually very large databases are needed for finding a model and to ensure a better generalization. But it's a quite tedious task to make the correct choice of the decay interval manually for all the subbands for estimating the RT for different RIRs. This makes the task virtually imposible to collect a large dataset with sufficient ground truth RTs for all the subbands.

Secondly, even if the Schroeder method fits (i.e., the frequency band is above Schroeder frequency and a suitable decay region is found for the subband), the task to compute the true T60 is still complex. Effects like ringing [10] ( artifacts that appear

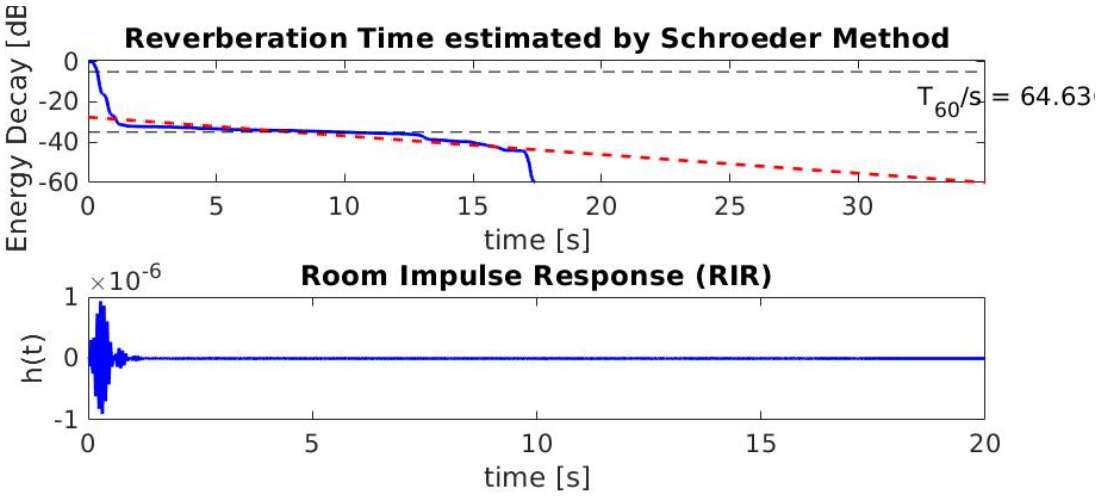


Figure 3.1: Visualization of sharp energy decay at the beginning of the time frame. The RIR is taken from ChurchMargaret RIR dataset. The subband corresponds to the 2nd subband with center frequency 31.6228 Hz using an octave filtebank.

as spurious signals near sharp transitions in a signal) and others show a very complex behaviour in the frequency subbands. With manual inspection in every subband, somehow it's possible to compute the true T60. But then this task will be very time consuming and virtually impossible for a very large dataset. An algorithm for automatic measurement of the true T60 using the Schroeder method is much needed for completing the first step towards a model based approach.

A normal iterative process involving different decay region could have easily solved the problem. Least square fitting is widely used as a state-of-the-art algorithm for determining the T60 using the Schroeder method. But the least square error (LSE) measurement doesn't give an optimal solution for choosing the best decay region in a first place. An analysis of the energy decay curve (EDC) calculated in the subband domain also shown that in some cases there are very sharp energy decays at the beginning of a time frame (shown in Fig. 3.1). To catch this characteristics, we need a very small decay region, which is also not possible with iterative algorithms.

## 3.1 Automatic Schroeder Method

An algorithm to estimate the subband RT automatically from a given RIR by means of the Schroeder method has been developed in this work. The general idea is to determine the RT for different decay regions and to take the RT with the lowest LSE,

which fulfills the threshold of maximum RT. A flow chart for this algorithm is shown in Fig. 3.2.

At the beginning of the iteration process, we set first the maximum threshold of the RT for a given RIR. We take the threshold as three times than the true fullband T60 for a particular RIR. We take three different decay regions for our observation, such as 10 dB, 20 dB and 30 dB. We start the iteration process with 30 dB. We calculate the T60s for different decay intervals with 30 dB region. Now we take the T60 corresponding to the lowest LSE. If this T60 fullfills the maximum RT threshold, then the iteration stops here with storing the T60 value and continue for next subband. Otherwise the iteration process continues with 20 dB and 10 dB region until the suitable T60 value for the specified subband is found.

As could be seen from the flow diagram, the minimum estimation occurs only after iteration over 10 dB decay region. In this step, if a candidate value for the T60 does not match, the minimum value of all T60 values is taken as final estimate.

## 3.2 Calculation of Subband RT Using Automatic Schroeder Method

We use the automatic Schroeder method discussed in the last section and create datasets for evaluating and investigating different model-based approaches for both DCT and octave filterbanks. We also visualize the ground truth subband RTs for different fullband ranges and analyse, if it is possible to create new parametric or learning-based models for the subband RT estimation.

Firstly, for this purpose we grouped the RIRs in different fullband T60 and direct-to-reverberant ratio (DRR) ranges. We used both DCT and octave filterbanks. For each of these filterbanks, we got a separate dataset. We termed this dataset as 'DCT full dataset' and 'Octave full datset'. In Fig. 3.3 and 3.4, we can see the results of the obtained T60s by the proposed automatic Schroeder algorithm for both the octave and DCT filterbank dataset.

From Fig. 3.3 and 3.4, it can be seen that the subband T60s are dependent much on the fullband T60, as the T60 in each subband for different fullband T60 ranges are somewhat homogeneous in nature.

It is also clear from the referenced figures that we can model subband T60s, if we discard all the outliers of the estimated RTs. To discard the outliers, we also need a measure for identifying the reliable and unreliable measurements. For this purpose, we

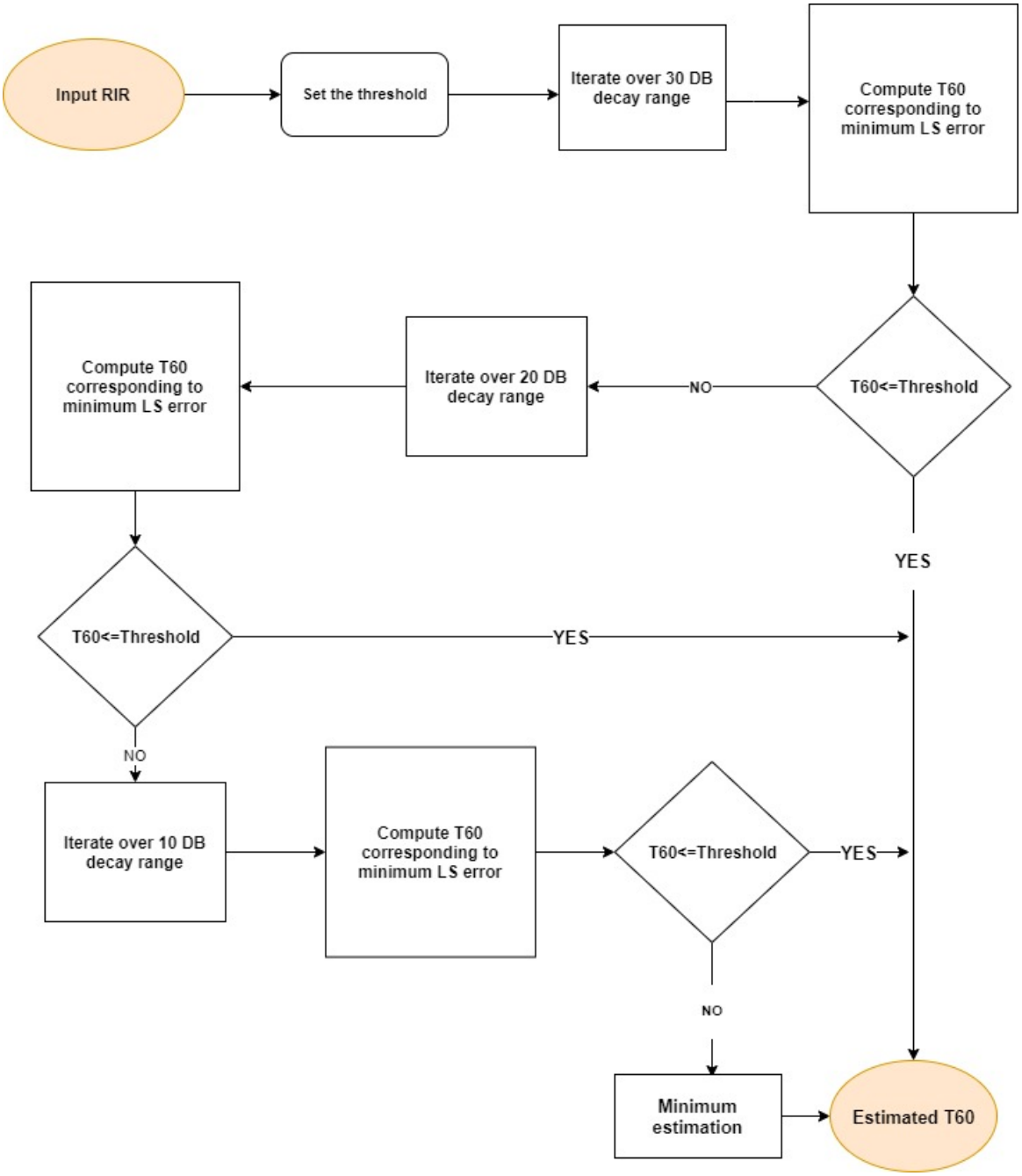


Figure 3.2: Flow diagram of the proposed algorithm for automatic RT estimation using the Schroeder method.

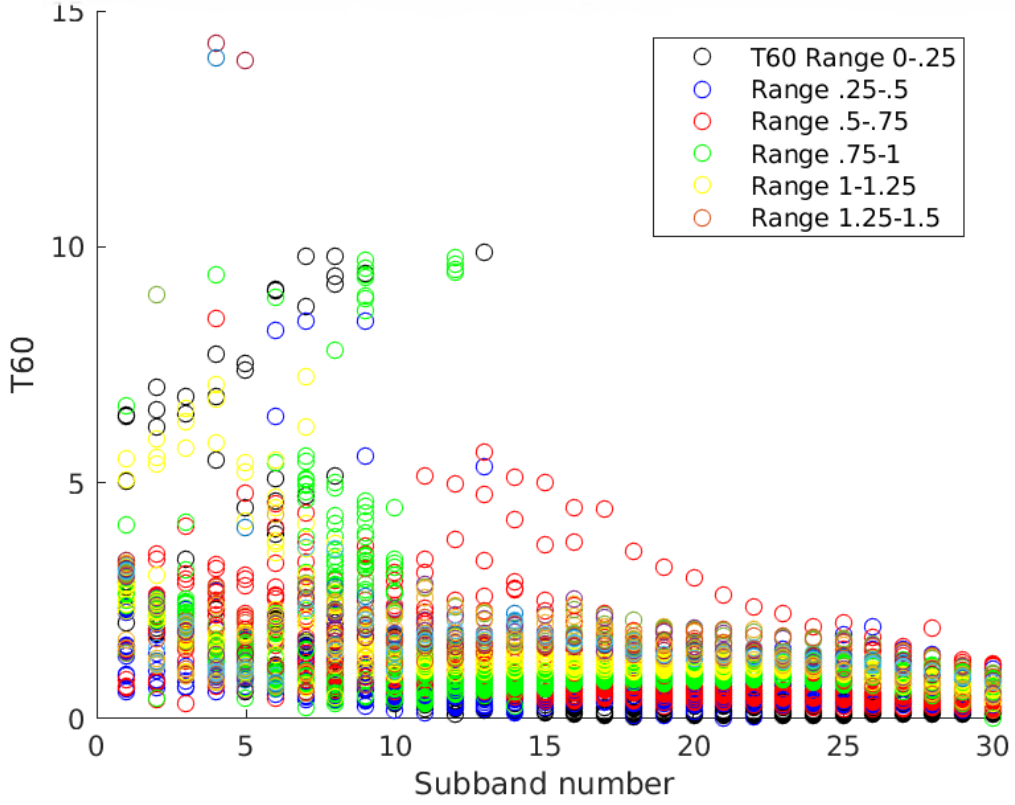


Figure 3.3: Computed T60s using the Schroeder method for the octave filterbanks.

created a reliable 'flag creator function' based on a threshold value. This reliable flag creator will analyse all the subband RTs for a given RIR and create a flag to indicate whether an estimate for a subband RT is reliable or not. If an estimate is flagged as unreliable, it is considered to be an outlier.

We divided all 30 subbands in 6 different regions, each consisting of 5 subbands. We set different thresholds for different subband regions. In case of a subband T60, the threshold will be based on the region from where that subband exists (i.e., for first five subbands, the T60 threshold will be the median T60 value of these subbands). We set a maximum and minimum threshold based on the median T60 and average LSE of that region. If an estimate fulfils both the constraints, we flagged that estimate as reliable, and as unreliable otherwise. In Fig. 3.5, the pseudo code for this process is shown.

So for developing a model-based approach, we need a clean dataset without any outlier RT estimations. In our investigation, we found that in using an octave filterbank resulted in at least one or more outliers RT estimations in most of the RIRs. We

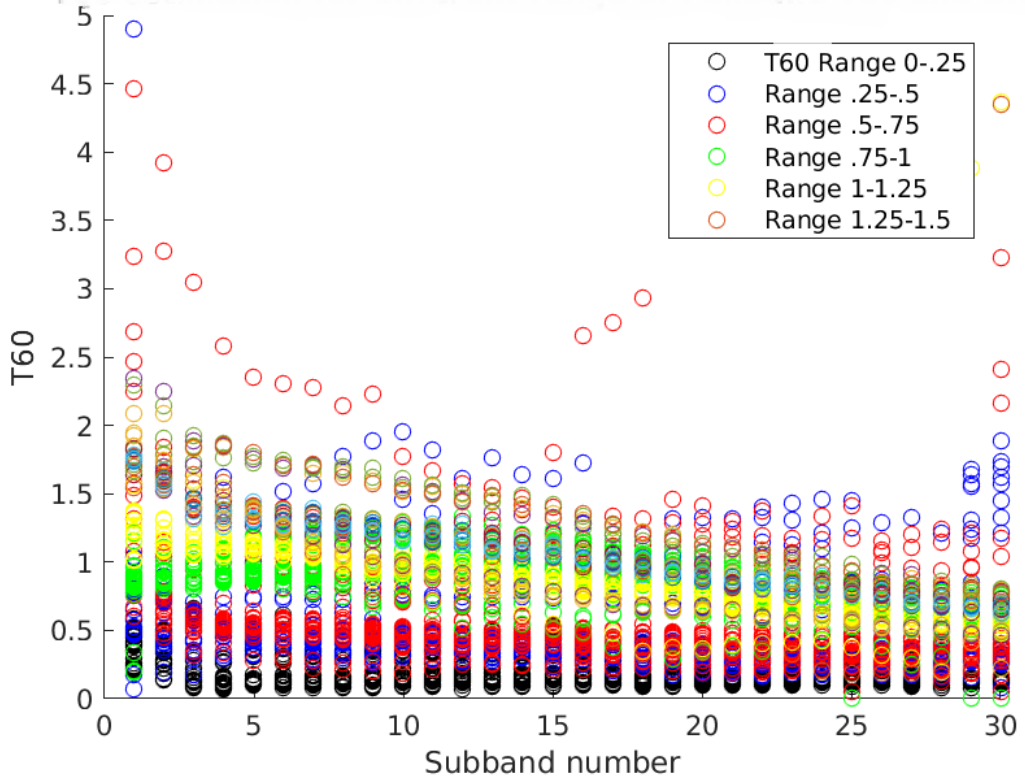


Figure 3.4: Computed T60s using the Schroeder method for the DCT filterbanks.

found only close to 60 RIRs, which contain no unreliable RT estimations. This makes it harder to create a polynomial-based model for an octave filterbank-based subband T60 estimation. As a consequence, we didn't consider an approach for checking the feasibility of an octave filterbank-based polynomial model for the subband T60 estimation. The result for the outlier analysis for the octave filterbanks is shown in the histogram in Fig. 3.6.

On the other side, using a DCT filterbank leads to better results. We get almost 370 clean RIRs without any outliers in any subbands. The result is shown in the histogram in Fig. 3.7.

This result also gives an indication towards the feasibility of a polynomial model-based approach for the T60 estimation in frequency subbands using a DCT filterbank.

In Fig. 3.8, the visualisation of the 'final clean dataset' for the DCT filterbanks is shown. It shows a clear trend for T60s in the frequency subbands. Compared to the Fig. 3.4, it's quite visible that there are no outliers or unreliable estimated subband T60s in this dataset. This result motivated us to consider a polynomial model-based

**Algorithm 1:** Pseudo Code for relaible flag creator

initialization;
Group the Subband RT in 6 groups;
$T60_{subband} = [T60_{group1}, T60_{group2}....., T60_{group6}]$;
Set threshold for each group;
$T60_{Threshold}$= median($T60_{group}$) ;
$LSE_{Threshold}$= mean($LSE_{group}$) ;
**for** *members* $n$ *in each group* **do**
  **if** $T60_{group}(n) \leq 1.5 \times T60_{Threshold}$ *and* $LSE(n) \leq 1.5 \times LSE_{Threshold}$ **then**
    Mark as reliable estimate;
  **else**
    Mark as unreliable estimate;
  **end**
**end**

Figure 3.5: Pseudo code for tagging an estimate as reliable or unreliable.

approach for estimating subband T60s using a DCT filterbank.

In the next chapter, we will evaluate and discuss different model-based approaches for subband T60 estimation using a DCT filterbank.

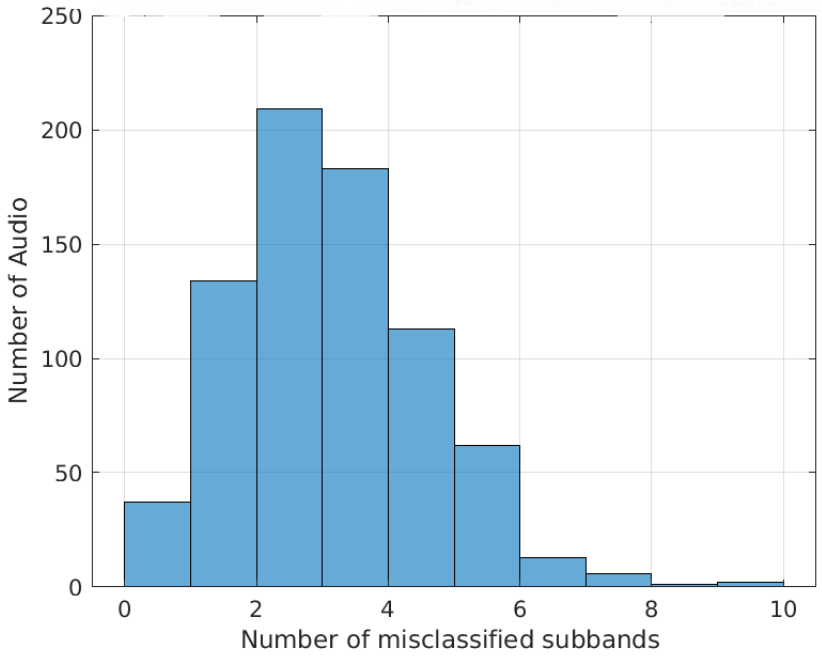


Figure 3.6: Histogram for RIR samples containing reliable and unrelaible subband RT estimations using an octave filterbank.

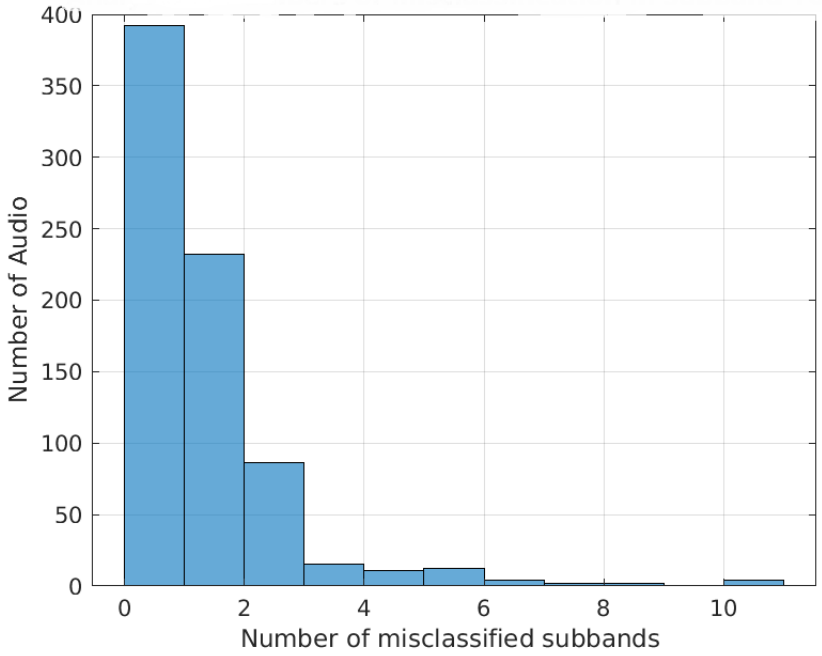


Figure 3.7: Histogram for RIR samples containing reliable and unrelaible subband RT estimations using a DCT filterbank.

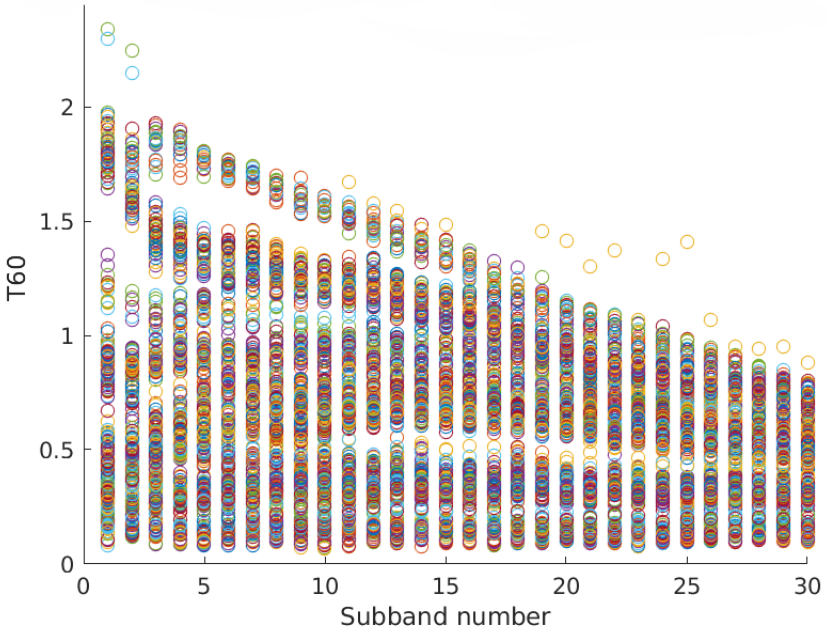


Figure 3.8: Visualisation of the clean dataset containg no ouliers or unrelaible subband RT estimates using a DCT filterbank.

# Chapter 4

# Model-Based Approaches

In this chapter, first we evaluated a model-based approach using DCT filterbanks discussed in [3]. Then we investigated two different model-based approaches for subband T60 estimation and a DNN model for fullband T60 estimation from the subband T60s. At the beginning, we worked on a first order polynomial fitting model. The observations and results of this model motivated us to make a DNN-based regression model for estimating fullband T60 from the subband estimations, and at the end we proposed another DNN-based regression model for estimating the subband T60s directly from a given RIR.

## 4.1 Marco Jeub's Model

In [3], two models for the frequency dependent RT are proposed which allow to calculate the RT in few subbands. A uniform DCT filterbank with 11 subbands is used in these models. The assumption is that a sufficient approximation of the frequency dependent RT can be obtained if the RT is known in two or three subbands. In general RT in these subbands are calculated using the Schroeder method. For the first case, the RT model named *RT Model 1* is given by the linear equation,

$$\hat{T}_{60}(f) = T_{60}(f_1) + \frac{T_{60}(f_2) - T_{60}(f_1)}{f_2 - f_1}(f - f_1). \tag{4.1}$$

Here $T_{60}(f_1)$ and $T_{60}(f_2)$ are the T60 at subbands with center frequencies $f_1$ and $f_2$. An extension of this model by including a third subband RT was also proposed, termed as *RT Model 2*,

$$\hat{T}_{60}(f) = \begin{cases} T_{60}(f_1) + \frac{T_{60}(f_2)-T_{60}(f_1)}{f_2-f_1}(f - f_1), & \text{for } 0 < f \leq f_2 \\ T_{60}(f_2) + \frac{T_{60}(f_3)-T_{60}(f_2)}{f_3-f_2}(f - f_3), & \text{for } f_2 < f \leq f_s/2 \end{cases}. \tag{4.2}$$

Here $T_{60}(f_3)$ denotes the RT in a third subband. The center frequencies are listed in

| Center frequency | $f_1$ | $f_2$ | $f_3$ |
|---|---|---|---|
| RT Model 1 | 0.8KHz | 7.2 KHz | - |
| RT Model 2 | 0.8KHz | 3.2 KHz | 7.2 KHz |

Table 4.1: Center frequencies for the frequency approximation models at $f_s$ =16 kHz.

Table 4.1.

### 4.1.1 Results

We evaluated the above mentioned models for our DCT full dataset. At first, we calculated the true subband RTs for 11 subands as described in [3], using the automatic Schroeder method discussed in Chapter 3. The average error according to (2.12) for *RT Model 1* was 0.0472s per subband and for *RT Model 2*, it was 0.1447s per subband. In Fig. 4.1 and 4.2 , the T60 over the subbands is plotted for two different RIRs. The red solid line gives the ground truth RT when the full dataset decompsed in total 11 subbands using a DCT filterbank. The dashed-dotted line and dashed lines with circle markers give the approximations with the proposed *RT Model 1* and *2*.

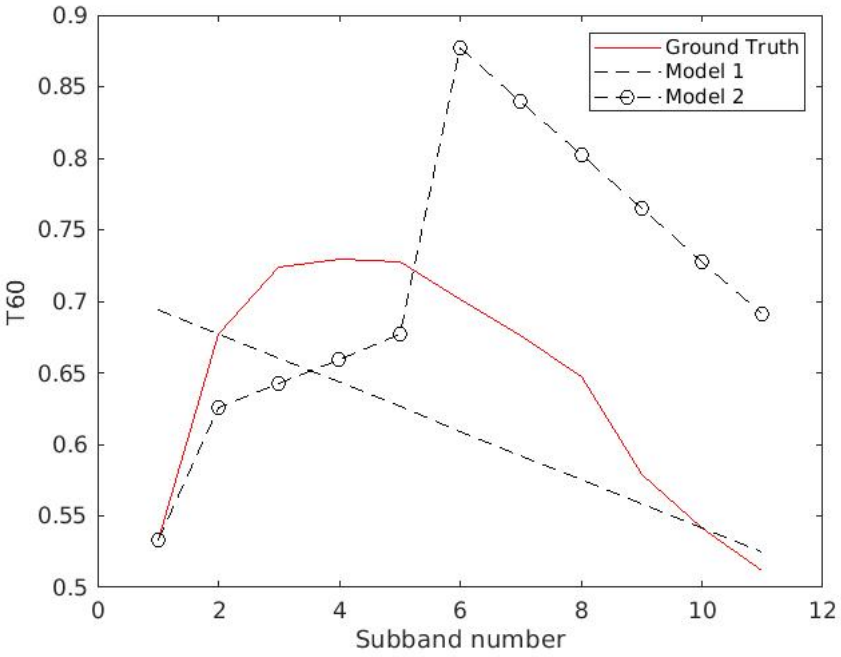


Figure 4.1: Example of subband T60 estimation for 11 subbands using Jeub's models (avearge error: *RT Model 1*= 0.0534s and *RT Model 2*= 0.1299s per subband).

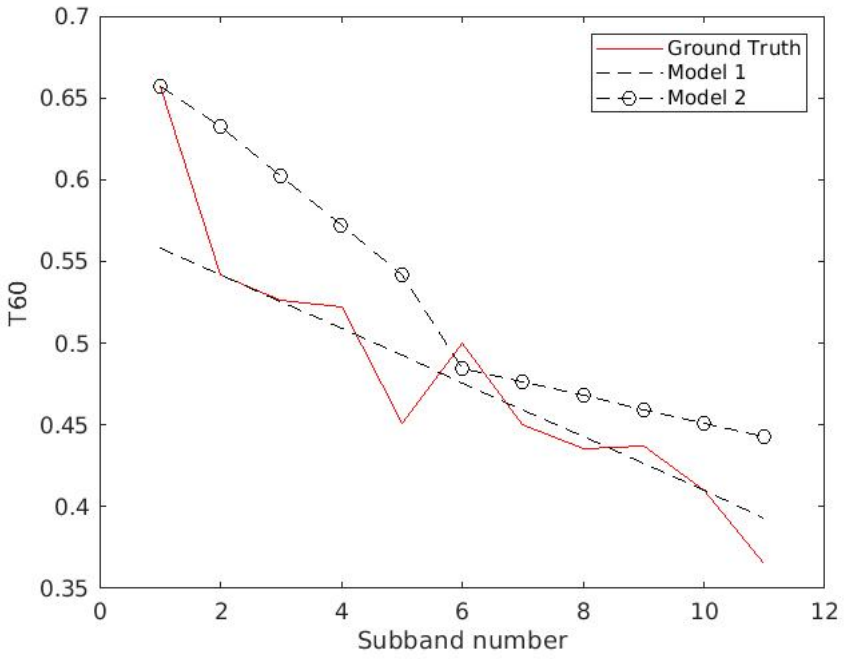


Figure 4.2: Example of subband T60 estimation for 11 subbands using Jeub's models (avearge error: *RT Model 1*= 0.0135s and *RT Model 2*= 0.0521s per subband).

In a second step, we evaluated these models using a DCT filterbank for 30 subbands. The ground truth subband RTs were calculated as before using the automatic Schroeder method. We used the same DCT full dataset. In this case, the average error according to (2.12) for *RT Model 1* was 0.0648s per subband and for *RT Model 2*, it was 0.1703s per subband. In Fig. 4.3 and 4.4, the T60 for all subbands is plotted for the same RIRs as shown in Fig. 4.1 and 4.2.

### 4.1.2 Observations

In previous section, the results of Jeub's model using a DCT filterbank for both, 11 and 30 subbands, are shown. But there is a serious problem in these models. These models totally depend on the assumption that the true T60s of two or three subbands are already known or can reliably be calculated using the Schroeder or other methods. If the RTs of these subbands are errorneous, the model will fail. An example for such a scenario is shown in Fig. 4.5. Here at 0.8 kHz center frequency, the subband RT was close to 4s, which is higher for all other subbands. It can be seen as an outlier, but as this center frequency band is used in the models, the estimated RTs are more

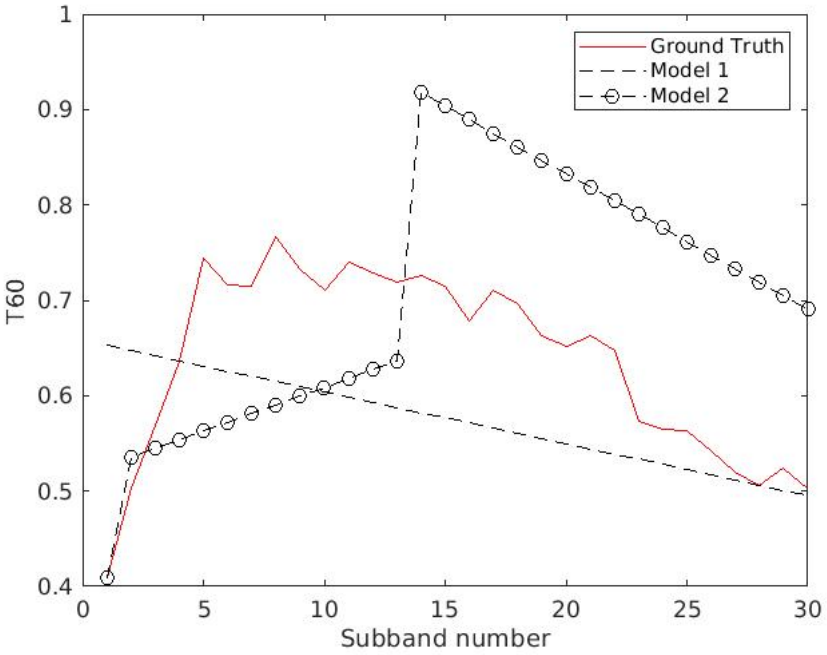


Figure 4.3: Example of subband T60 estimation for 30 subbands using Jeub's models (avearge error: *RT Model 1*= 0.0895s and *RT Model 2*= 0.1565s per subband).

erroneous. The average error for *RT Model 1* was 1.1812s per subband and for *RT Model 2*, it was 1.0642s per subband.

## 4.2 First Order Polynomial Model

As shown earlier, the subband T60s largely depend on the fullband T60 and, therefore, we grouped our dataset on different fullband T60 ranges (state in steps of 100ms starting with 0.1s). Then we calculated two first order polynomial models for this dataset.

For the first model, we tried to fit the data for all the subbands. Considering the Schroeder frequency, in the second model we fitted the data only for the upper subbands starting from the 15th subband. As an uniform DCT filterbank with 30 subbands was used, the 15th subband corresponding to the quarter of the sampling frequency (i.e., 4 kHz for the considered sampling frequency of 16 kHz), which is well above the Schroeder frequency in most cases. In Chapter 2, it was dicussed based on common model-based approaches that there is an assumption of a monomaniacal decay of the RT in higher subbands. Considering all these aspects, we take the 15th

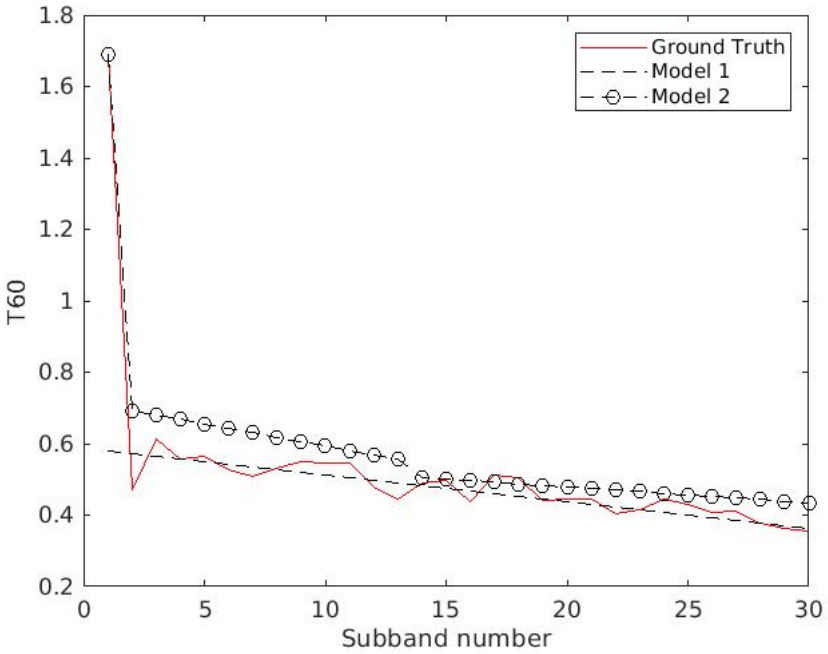


Figure 4.4: Example of subband T60 estimation for 30 subbands using Jeub's models (avearge error: *RT Model 1*= 0.0249s and *RT Model 2*= 0.0632s per subband).

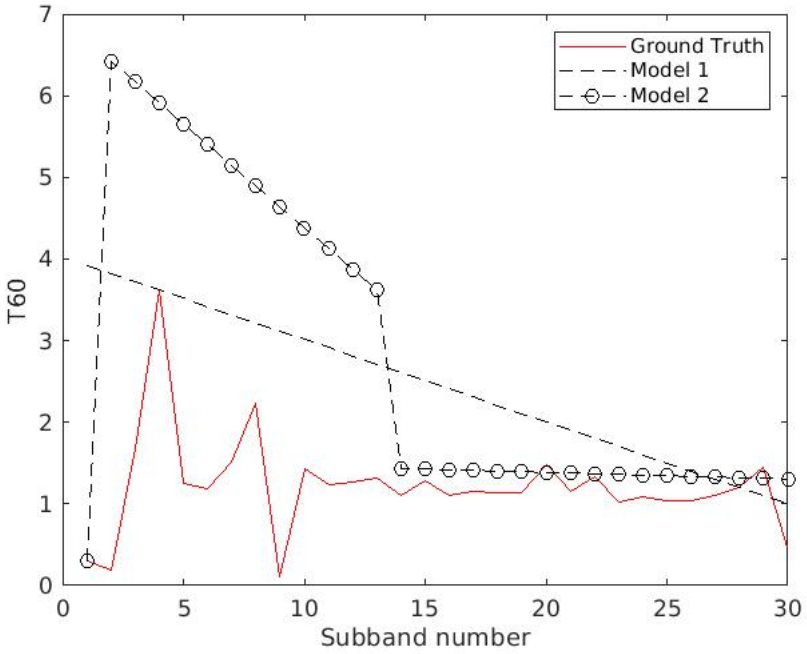


Figure 4.5: Complication of Jeub's models for the subband T60 estimation.

subband as the reference subband in our second model.

For making this polynomial model, we used the Matlab function *polyfit*. $Polyfit(x, y, n)$ returns the coefficients for a polynomial $p(x)$ of degree $n$ (in our case $n = 1$) that is a best fit in a least-squares sense for the data in $y$.

$$p(x) = p_1 x^n + p_2 x^{n-1} + ..... + p_n x + p_{n+1}. \tag{4.3}$$

Here in (4.3), the coefficients in $p$ are in descending powers, and the length of $p$ is $n + 1$.

### 4.2.1 Results

As expected from the results shown in Chapter 2, the second model for the upper subbands fitted well with a first order polynomial model. The results of the described approach are shown in Fig 4.6.

It can be observed that the regression line for the model is represented with the black line and the circles are the ground truth T60 in the subbands obtained with the automatic Schroeder method described in the previous chapter. This result also re-verifies the findings of the Hybrid model described earlier, but in this case we used a DCT filterbank. The overall mean absolute average error considering all the T60 ranges was calculated using (2.12) and the following equation,

$$mean_{AE} = \frac{\sum_{N=1}^{N} \frac{\sum_{N_s=1}^{N_s} AE}{N_s}}{N}. \tag{4.4}$$

This model achieves 0.0400s error per subband. Here in (4.4), $N$ represents total number of T60 ranges we considered, $N_s$ represents the number of RIRs in every T60 range and AE is the average absolute error calculated using (2.12) for a given RIR.

Unfortunately, the polynomial model doesn't work well for the first case where we tried to fit the model for all the subbands. The findings are shown in Fig 4.7. A first order polynomial model cannot approximate well the T60s in the lower subbands. The overall mean absolute average error calculated using (4.4), in this case was 0.1010s per subband, which shows that a general model for all subbands is not found by polynomial fitting.

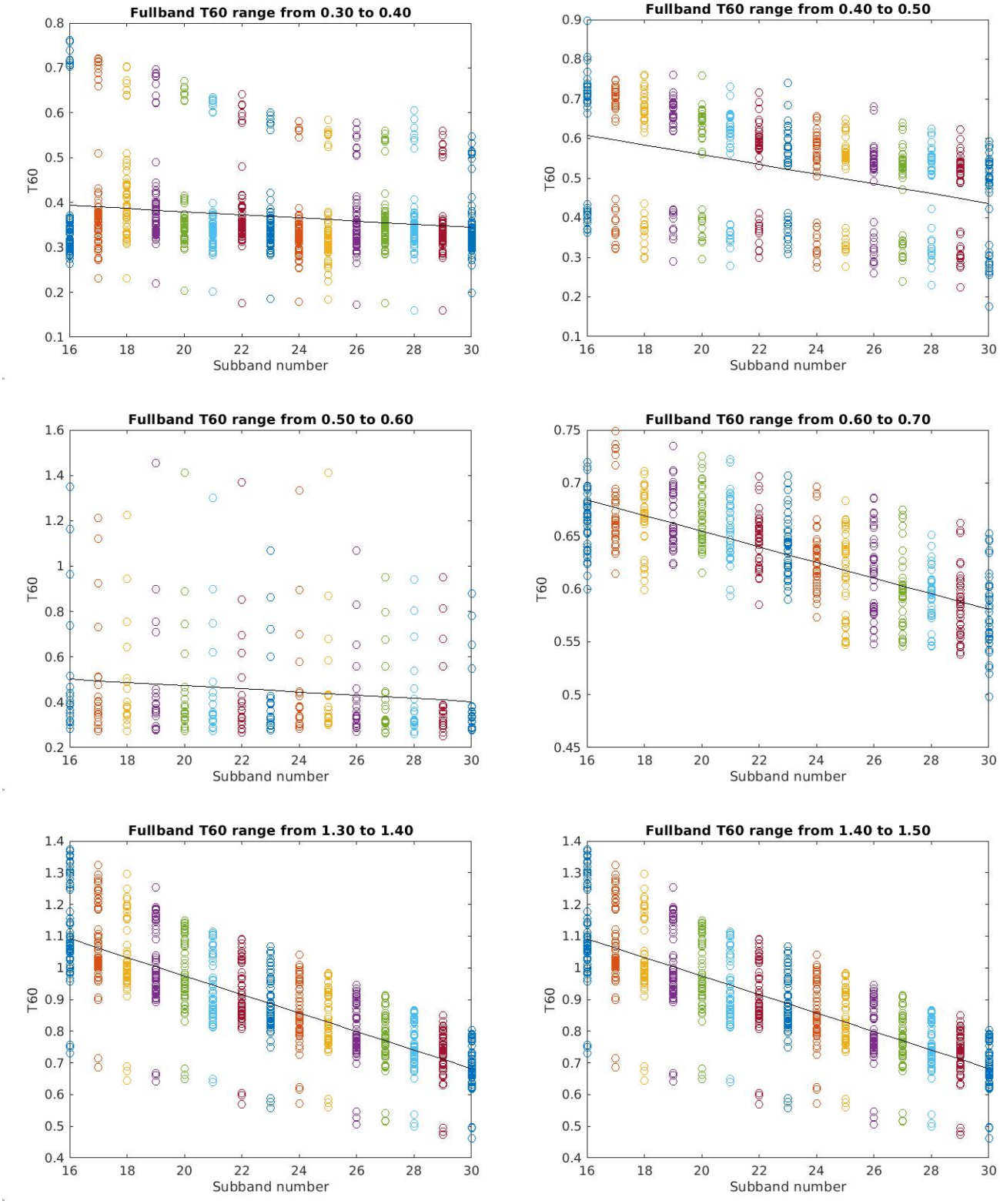


Figure 4.6: Examples of polynomial regression model for upper subbands for different fullband T60 ranges.

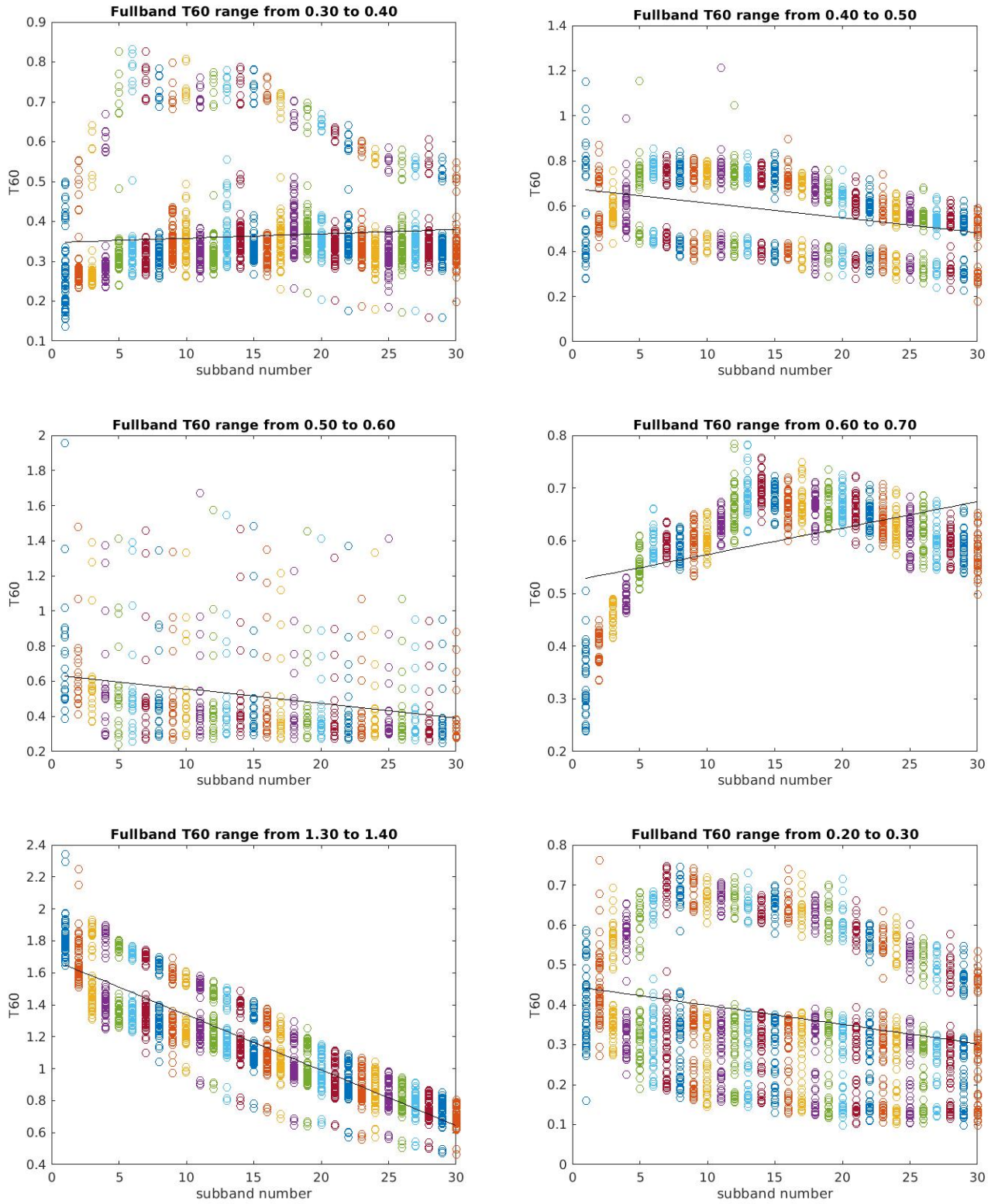


Figure 4.7: Examples of polynomial regression model for all subbands for different fullband T60 ranges.

### 4.2.2 Observations

As we have seen in the previous results, a model-based approach work well for the upper subbands. Apart from that there are some other important observations. The model always works well for both cases (i.e., all subbands and just the upper subbands) for higher fullband T60 ranges. For the model considering all subbands, the average error up to fullband T60 range 0.50s (first four groups) was 0.1115s, whereas for the rest of the fullband T60 ranges (last seven groups ) till 1.50s the average error was 0.0922s. Same observation has been made for the upper subband model. The average error up to fullband T60 range 0.50s was 0.049s and for the rest of the fullband T60 ranges, the average error was 0.0319s. This shows an indication of the limitation for computing T60 in the subbands using the Schroeder method for low fullband T60 ranges. Along with it, we observed different subbands T60 patterns even in the same fullband T60 ranges. This emphasizes the dependency of the individual room characteristics on the subband T60s. These two observations show that it's quite impossible to make an universal model for the subbands T60 estimation for different fullband ranges and room characteristics using the analytical polynomial fitting model.

For the results, we have seen that there exist some patterns in the subband T60s in different fullband T60 ranges. This inspired us to investigate whether there is any relation between the subband T60s and the fullband T60. It was later investigated that there might exist some relationship between these two quantities.

To examine this observation, an analytical model to estimate the fullband T60 from the subband T60s has been formulated as follows,

$$RT_{T60} = \frac{\frac{\sum_{n=1}^{30} sub_{T60}(n)}{30} + \frac{\sum_{n=1}^{10} sub_{T60}(n)}{10} + \frac{\sum_{n=1}^{5} model_{T60(n)}}{5}}{3} - error_{model}. \tag{4.5}$$

Here $RT_{60}$ is the estimated fullband T60 from the subband T60s and $sub_{T60}$ represents the subband T60s for a given RIR. The first summation of $sub_{T60}$ for all the subbands (here 30 subbands) provides an average weighting. The motivation behind the second summation of $sub_{T60}$ upto subband 10 was to give more weights to the lower subbands. The $model_{T60}$ provides the regression values for that range, which is obtained from the previous first order polynomial model. This $model_{T60}$ represents the regression line, which has been shown in Fig 4.7 . And finally we subtract the model error $error_{model}$. This is the average error (calculated using (2.12)) for the given RIRs, obtained with the previously discussed first order polynomial model.

The result was quite promising. The average error for the fullband T60 ranges from

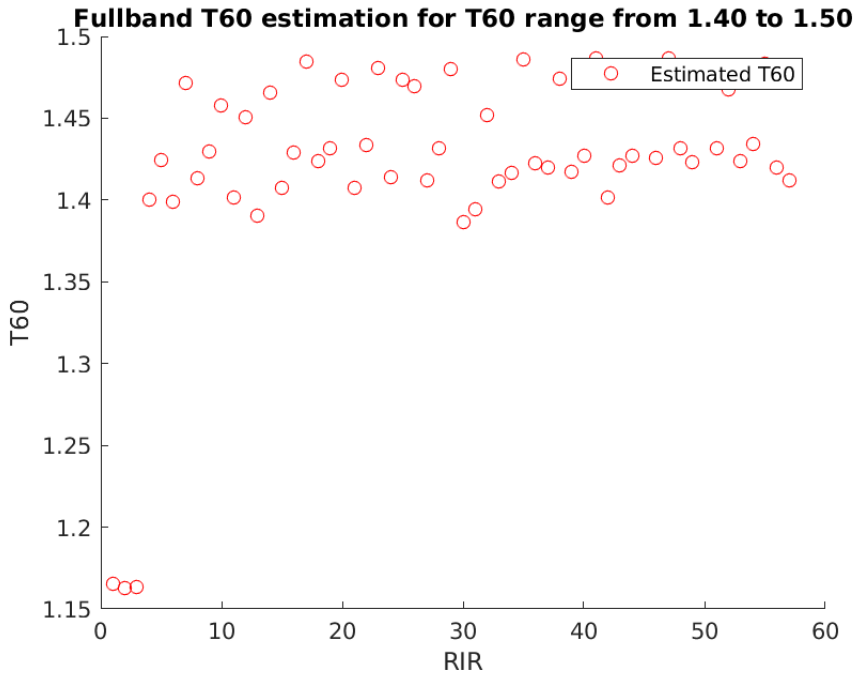


Figure 4.8: Fullband T60 estimation from the subband RTs.

1.4s to 1.5s was 0.0485s (see Fig 4.8). This result has motivated us to make a DNN model for estimating the fullband T60 from the subband T60s.

## 4.3 DNN Model for the Fullband T60 Using the Subband Estimations

As described already in the previous section, it's possible to estimate the fullband T60 from the subband estimates. Some efforts have been made to use the subband knowledge for blind estimation of the fullband T60. In [4], a method is described to estimate the fullband T60 exploiting the subband's information by a maximum likelihood method. Here a convolutional neural network for estimating the fullband T60 from subband T60s is used. The network architecture is shown in Fig. 4.9.

At first we used the clean dataset without any outliers, as presented in Chapter 3. We separate the dataset into two completely separate training and test datasets using the open source Scikit-learn library [11]. Here we got 253 samples for training and 109 samples for testing. We normalized the training data in between (0-1) using MinMaxScaler function provided by the Scikit-learn library.

In another test case for checking the generalization of our model, we used the dev

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
gaussian_noise_1 (GaussianNo (None, 30, 1)             0
_________________________________________________________________
conv1d_1 (Conv1D)            (None, 30, 32)            128
_________________________________________________________________
max_pooling1d_1 (MaxPooling1 (None, 10, 32)            0
_________________________________________________________________
flatten_1 (Flatten)          (None, 320)               0
_________________________________________________________________
dense_1 (Dense)              (None, 512)               164352
_________________________________________________________________
dropout_1 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_2 (Dense)              (None, 32)                16416
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 33
=================================================================
Total params: 180,929
Trainable params: 180,929
Non-trainable params: 0
_________________________________________________________________
```

Figure 4.9: Network architecture for the fullband T60 estimation.

dataset of the ACE Challenge which contains 14 RIRs. The rest of the clean dataset (all the samples of clean dataset except the samples from ACE challenge dataset) was used for training. Data processing was also same in this case (i.e., normalization, etc.).

After achieving satisfactory result in that case, we trained our model with the full dataset containing outliers too. We seprated 70 % data for training and 30 % for test. We did it for both, the DCT filterbank dataset and octvae filterbank dataset.

### 4.3.1 Results

In our experiment, the model was evaluated using the mean square error and mean absolute error loss.

$$MSE = \frac{\sum_{i=1}^{N}(y(i) - \hat{y}(i))^2}{N} \tag{4.6}$$

$$MAE = \frac{\sum_{i=1}^{N} \mid (y(i) - \hat{y}(i)) \mid}{N} \tag{4.7}$$

Here $y(i)$ is the ground truth RT and $\hat{y}(i)$ is the estimated RT for a given RIR and $N$ is the total number of training or test samples.

In the first case, where we used only the clean dataset, during testing we achieved 0.0557s mean absolute error and 0.0048s mean square error. The results of testing for 92 randomly chosen samples are shown in Fig. 4.10.

In test case with only the ACE challenge dataset, the MAE was 0.0662s and MSE was 0.0069s. The result is shown in Fig. 4.11.

The result with the full dataset containing outliers was also satisfactory. For the DCT filterbanks, the MAE was 0.058s and MSE was 0.0027s (test samples were randomly chosen). The predicted result is visualized in Fig. 4.12.

The subband T60 estimations resulted with more outliers using an octave filterbank, as discussed in the Chapter 3. Surprisingly, still in case of the octave filterbanks dataset, the test result was also good. The MAE was 0.0609s and MSE was 0.0062s . In Fig. 4.13, we can see the results with the octave filterbanks dataset.

The results and important parameters for fullband RT estimation from the subband RTs for different datasets are listed in the following Table 4.2.

### 4.3.2 Observations

It is clear from the results that a DNN model can provide a good estimate of the fullband T60 from the subband RT estimates. The generalization of this method has

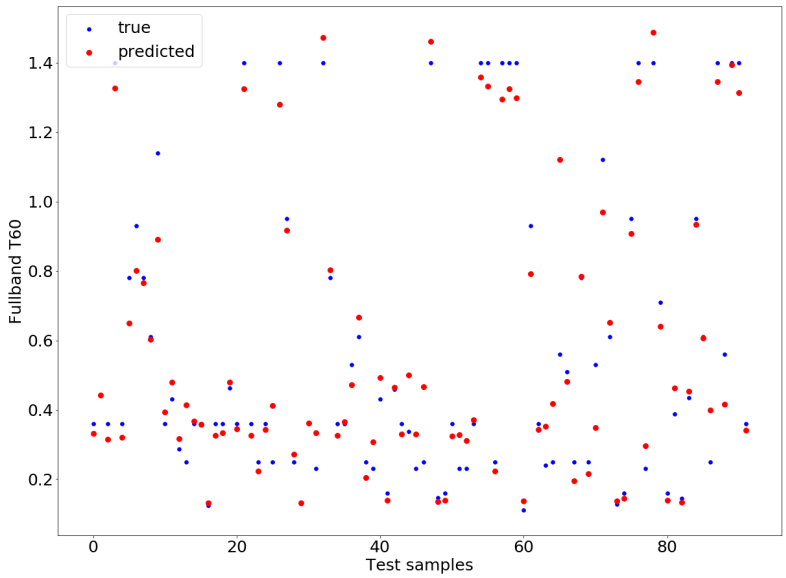


Figure 4.10: Fullband T60 estimation for the clean dataset using the DCT filterbanks.

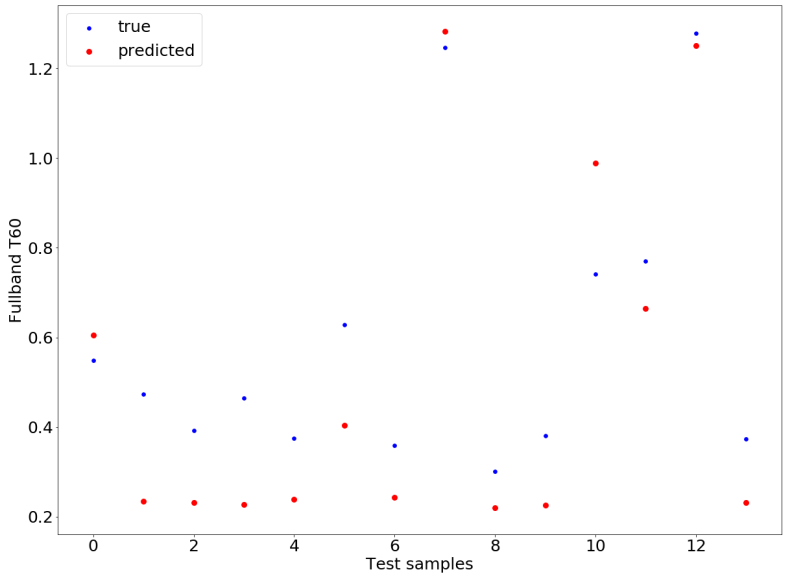


Figure 4.11: Fullband T60 estimation for the ACE dataset using the DCT filterbanks.

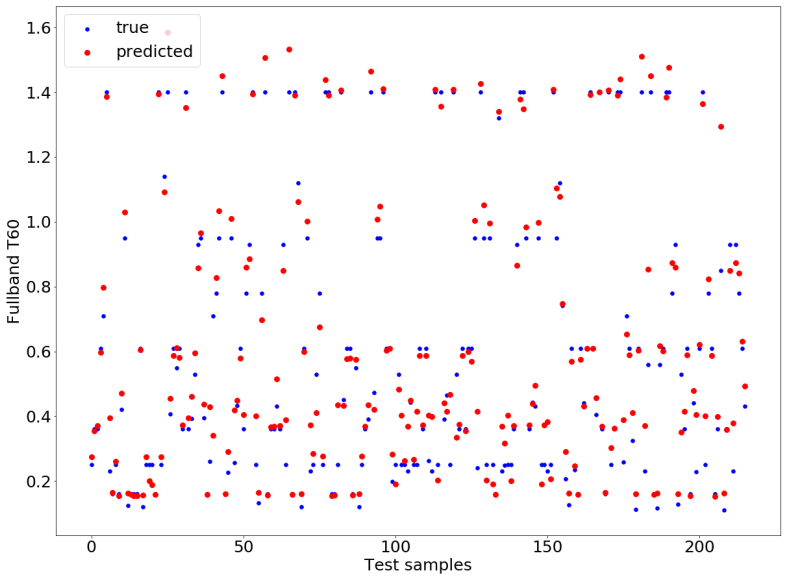


Figure 4.12: Fullband T60 estimation for the full dataset using the DCT filterbanks.

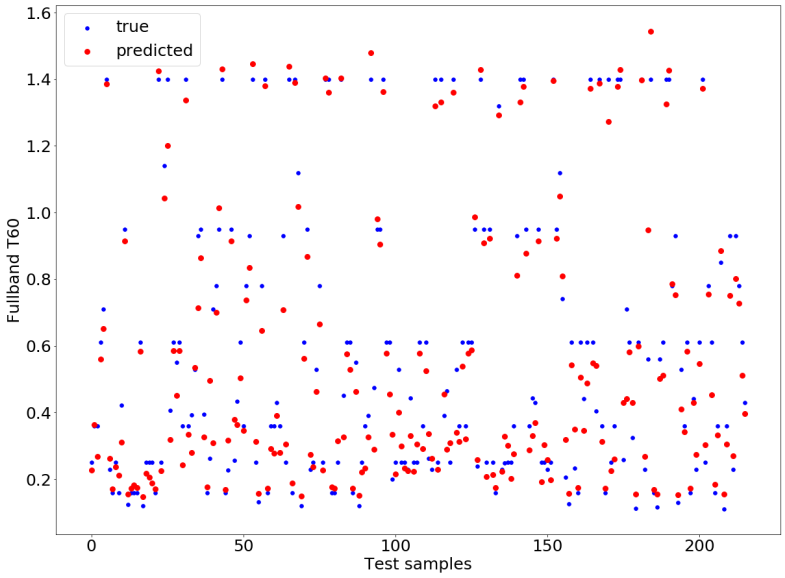


Figure 4.13: Fullband T60 estimation for the full dataset using the octave filterbanks.

| Dataset Name | Clean Dataset | ACE Dataset | DCT Full Dataset | Octave Full Dataset |
|---|---|---|---|---|
| Training Samples | 253 | 348 | 532 | 532 |
| Test Samples | 109 | 14 | 216 | 216 |
| Filterbanks | DCT | DCT | DCT | Octave |
| outliers | No | No | Yes | Yes |
| MAE | 0.0557 | 0.0662 | 0.058 | 0.0609 |
| MSE | 0.0048 | 0.0069 | 0.0027 | 0.0062 |

Table 4.2: Results and parameters for the fullband T60 estimation.

been also proved by the ACE challenge test dataset. One of the important observations is that this model can be used for both, DCT filterbanks and octave filterbanks. Finally the most important one is that the result with the dataset which contain many outliers, was also satisfactory. This indicates that we can use this model for the fullband RT estimation on top of a blind RT estimation for the subbands. Sufficient training data is needed to ensure a sufficient generalization for various acoustic enclosures.

## 4.4 DNN Model for the Subband T60 Estimation

In 4.2, it was shown that with analytical polynomial fitting, a universal model cannot be realized. Now a CNN-based regression model for estimating the subband T60s directly from a given RIR is proposed. Previously, different approaches have been presented for using a DNN model for estimating the fullband T60 directly from a given RIR using the DNN architecture. One of such approaches is discussed in [12].

In our model for a given RIR, we extract the spectogram feature. We used the already discussed clean dataset. The spectrogram is calculated for time frame of 2s. First we take the STFT of a given frame with $nfft = 512$ (frequency bins), window length was same as $nfft$ length and the hop length was 128. We used the open source Librosa library [13] for calculating the STFT. After that we take the absolute value of it and square it for having a power spectogram. After these feature extraction processes, we get a data array of size $257 \times 251$ for every 2 seconds duration frame. This is our input to the CNN. In total we get 1143 samples from the clean dataset, which are in total 362 RIRs. Among these 800 samples (i.e., a data array for 2s duration of a RIR) was used for training and 343 samples were separated for testing. For splitting the dataset into training and test dataset the Sckit-learn library [11] was used as before. The network details are shown in Fig. 4.14.

### 4.4.1 Results

To evaluate this model, we obtained the ground truth value of the subband T60s by the automatic Schroeder method discussed in Chapter 3. The result with this CNN model is quite promising for subband T60 estimation. The MAE of (4.7) was 0.1207s and the MSE of (4.6) was 0.0320s. The loss plot is shown in Fig. 4.15. The results for a random RIR is also shown in Fig. 4.16.

The results and important parameters for the subband RT estimation from a given RIR for the DCT filterbanks full dataset are listed in Table 4.3.

### 4.4.2 Observations

The result from this CNN model shows that it's possible to estimate the subband T60s directly from a given RIR. A more sophisticated feature extraction and CNN model might lead to a more accurate estimation. One of the important findings from our model is that it approximates well for the higher fullband T60 than the lower one. The same problem we encountered in our polynomial model too. The reason behind

| Layer (type) | Output Shape | Param # |
|---|---|---|
| gaussian_noise_1 (GaussianNo | (None, 257, 251, 1) | 0 |
| conv2d_1 (Conv2D) | (None, 257, 251, 64) | 640 |
| max_pooling2d_1 (MaxPooling2 | (None, 85, 83, 64) | 0 |
| conv2d_2 (Conv2D) | (None, 85, 83, 32) | 18464 |
| max_pooling2d_2 (MaxPooling2 | (None, 28, 27, 32) | 0 |
| conv2d_3 (Conv2D) | (None, 28, 27, 32) | 9248 |
| max_pooling2d_3 (MaxPooling2 | (None, 9, 9, 32) | 0 |
| flatten_1 (Flatten) | (None, 2592) | 0 |
| dense_1 (Dense) | (None, 1024) | 2655232 |
| dropout_1 (Dropout) | (None, 1024) | 0 |
| dense_2 (Dense) | (None, 128) | 131200 |
| dense_3 (Dense) | (None, 30) | 3870 |

Total params: 2,818,654
Trainable params: 2,818,654
Non-trainable params: 0

Figure 4.14: Network architecture for the subband T60 estimations.

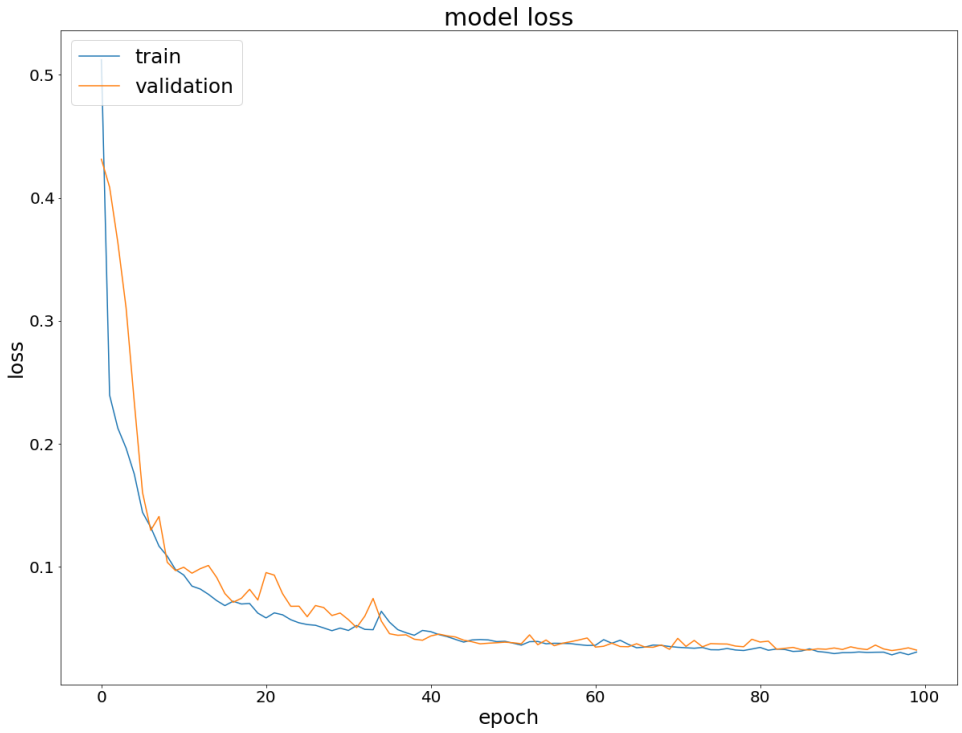


Figure 4.15: Training MSE Loss plot for the subband T60 estimations.

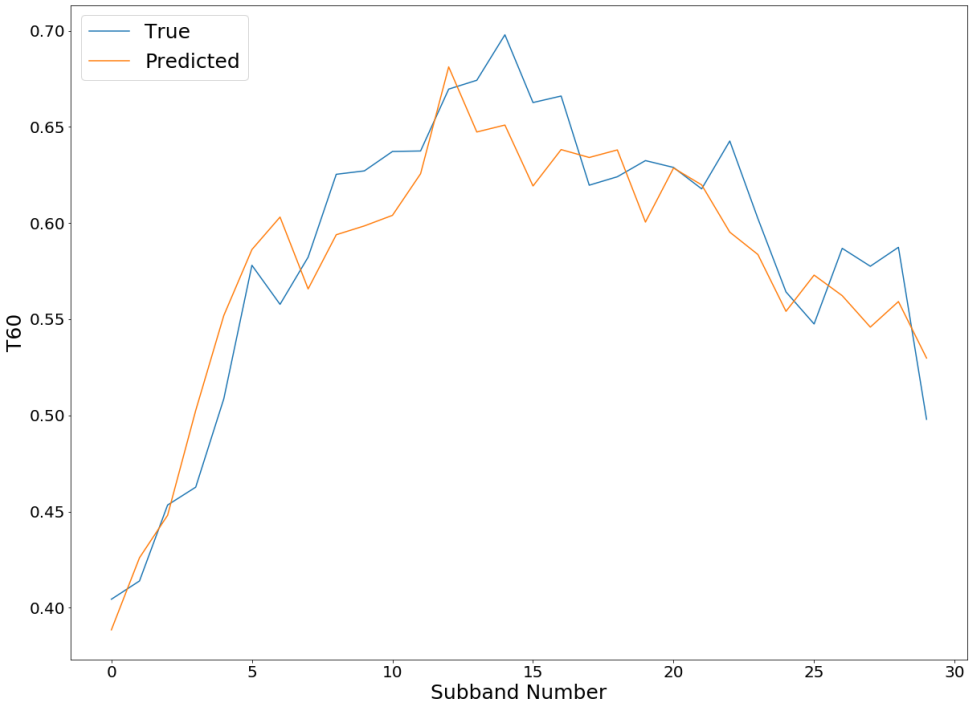


Figure 4.16: Subband T60 estimations using CNN model.

| Dataset Name | DCT full dataset |
|---|---|
| Input features | Spectogram of 2s sequence of given RIR |
| Input shape | 257×251 |
| Training Samples | 800 |
| Test Samples | 343 |
| Filterbanks | DCT |
| outliers | No |
| MAE | 0.1207 |
| MSE | 0.0320 |

Table 4.3: Results and parameters for the subband T60 estimation.

it might be that the baseline Schroeder method itself can not measure the T60 well in each subband.

# Chapter 5

# Conclusion

In this research work, various methods and models for estimating the subband and fullband T60 were discussed. It has been shown that there is a pattern for the subband T60s. In a step to investigate different model-based subband T60 estimators, an automatic Schroeder method is proposed. A subject for further work is to investigate this method for a large database with RIRs containing ground-truth values for the subband T60s.

A polynomial-based fitting model was also developed using a DCT filterbank. However, this method cannot universally approximate well the subband T60s for all the subbands. In contrast, this polynomial model shows better results for the upper subbands. Perhaps this type of model can be used on top of other blind estimation algorithms. Two models from [3] using a DCT filterbank were also evaluated. These models show quite satisfactory results for different RIRs (one achieves 0.05s average error per subband), but it assumes that the T60 of two or three subbands are already known or can be estimated reliably. This assumption makes it quite difficult to use this method for blind estimation of the subband T60, where the anchor T60 values can be erroneous..

Using a DNN model can help to find a model for the subband T60 estimation for different RIRs. A CNN-based model was developed to estimate the subband RTs blindly from a given RIR. This model achieves 0.12s mean absolute error. Notably our dataset was small, containing only about 350 RIRs. A big dataset with ground-truth suband RTs might lead to a better model.

The investigation for fullband T60 estimation from subband T60s shows promising results. A 1-D CNN model is also proposed for the estimation of the fullaband T60 from the subband estimates. This model achieves 0.05s MAE by using the DCT filterbank dataset, whereas for the octave filterbank the MAE was 0.06s. This kind of model can be used very easily for fullband T60 estimation due to it's simple structure and low computational complexity. It can be used on top of any model for blind estimation of subband T60s. Further investigations with larger, labeled datasets are needed to

further verify the developed models.

# List of Figures

# Bibliography


[1] S. Li, R. Schlieper, and J. Peissig, "A hybrid method for blind estimation of frequency dependent reverberation time using speech signals," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, May 2019.

[2] H. W. Löllmann, A. Brendel, W. Kellermann, and P. Vary, "Single-channel maximum-likelihood T60 estimation exploiting subband information," *Proceedings of the ACE Challenge Workshop*, pp. 1–5, October 2015.

[3] M. Jeub, "Joint dreverberation and noise reduction for binaural hearing aids and mobile phone," *PhD thesis, published at RWTH Aachen University*, 2012.

[4] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An improved algorithm for blind reverberation time estimation," *In Proceedings of International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 1–4, 2010.

[5] H. W. Löllmann and P. Vary, "Estimation of the reverberation time in noisy environments," *in Proccedings International Workshop Acoustic Echo Noise Control (IWAENC)*, pp. 1–4, 2008.

[6] J. Eaton, A. H. Moore, N. D. Gaubitch, and P. A. Naylor, "The ACE challenge - corpus description and performance evaluation," *in Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 15.

[7] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, no. 2, p. 409–412, 1965.

[8] S. M. Khezri and P. J. Shalkouhi, "The Schroeder Frequency of Furnished and Unfurnished Spaces," *Graduate School of the Environment and Energy, Science and Research Branch, Islamic Azad University, Tehran, Iran*, pp. 1–4, January 2012.

[9] W. Sabine, "Collected Papers on Acoustics," *Harvard University*, 1922.

[10] "Ringing artifacts," *Wikipedia, The Free Encyclopedia.*, 2019.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[12] M. Lee and J. H. Chang, "Deep Neural Network Based Blind Estimation of Reverberation Time Based on Multi-channel Microphones," *Acta Acustica United with Acustica*, vol. 104, p. 486 – 495, 2018.

[13] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in python," in *Proceedings of the 14th python in science conference*, vol. 8, 2015.