ESTIMATION OF THE FREQUENCY DEPENDENT REVERBERATION TIME BY MEANS OF WARPED FILTER-BANKS

Heinrich W. Löllmann and Peter Vary

Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, 52056 Aachen, Germany
{loellmann,vary}@ind.rwth-aachen.de

ABSTRACT

An improved approach for the estimation of the frequency dependent reverberation time (RT) by means of allpass transformed filter-banks is presented. It is shown that by means of these warped filter-banks, a much more accurate RT estimation at lower frequencies can be obtained than by octave filter-banks, which are commonly used for the estimation of the frequency dependent RT. Furthermore, allpass transformed filter-banks can achieve a much better approximation of the non-uniform frequency resolution of the human auditory system than octave filter-banks. A uniform or non-uniform (auditory) frequency resolution can thereby be simply adjusted by a single allpass coefficient.
The RT estimation can be done with an allpass transformed DFT or DCT filter-bank. The warped DCT filter-bank is of special interest as it provides real-valued subband signals. This facilitates the use of a maximum-likelihood (ML) estimator for either a non-blind estimation of the frequency dependent RT from a room impulse response or a blind estimation from a reverberant speech signal.

Index Terms— reverberation time, frequency warping, frequency dependent decay, sound decay measurement

1. INTRODUCTION

The *reverberation time* (RT) T_{60} is one of the most important quantities in room acoustics and plays a crucial role in the evaluation of enclosed auditory spaces such as lecture rooms or concert halls [1]. Furthermore, knowledge about the RT can also be exploited for speech dereverberation [2, 3].

The RT is defined as the time interval in which the energy of a steady-state sound field decays 60 dB below its initial level after switching off the exciting sound source. This time interval can be calculated either by the ensemble average of different sound decays or from a measured *room impulse response* (RIR) by means of the Schroeder method [4, 5].

The sound decay or RT is often measured within different frequency bands to take the frequency dependent sound absorption at a surface into account. A well-established approach to determine the frequency dependent RT is to filter the RIR by bandpass filters and to apply the Schroeder method in each subband. The used analysis filters have either full-octave bands or 1/3-octave bands according to [6]. Such octave filters constitute a so-called constant-Q filter-bank and account for the non-uniform frequency resolution of the human auditory system, e.g., [7].

A known problem of this approach is that the octave bandpass filters for the lower frequencies have a very small bandwidth, which leads to unreliable RT estimates at these frequencies. In [8], it is recommended that the product of bandwidth and RT should exceed a value of 16 to obtain reliable results. An approach to alleviate this problem is to use a wavelet filter-bank with 1/3-octave bands [9]. Another approach is to perform a so-called time-reversed decay measurement by means of zero-phase bandpass filters [10].

The use of bandpass filters with a very low bandwidth can be avoided by employing a uniform filter-bank for the calculation of the frequency dependent RT. One possibility is the use of a DFT filter-bank. The RT can be calculated from the complex subband signals by means of the energy decay relief (EDR) [11]. However, such an approach is not comparable with the use of an octave filter-bank as it does not account for the non-uniform frequency resolution of the human auditory system.

In this contribution, it is shown that allpass transformed filter-banks are very attractive for the calculation of the frequency dependent RT. These warped filter-banks account for the non-uniform frequency resolution of the human ear more accurately than octave filter-banks and provide more reliable RT values at low frequencies.

This paper is organized as follows: In Sec. 2, the design of the proposed filter-banks is introduced and compared to that of an octave filter-bank. The estimation of the frequency dependent RT by means of different non-uniform filter-banks is elaborated in Sec. 3. A benefit of the proposed warped DCT filter-bank is that common maximum-likelihood (ML) based techniques to estimate the RT can be applied, which is investigated in Sec. 4. The paper concludes with a summary by Sec. 5.

2. WARPED FILTER-BANKS

The *allpass transformation* is a well-known technique to design a digital filter-bank with a non-uniform time-frequency resolution [12–14]. In the process, the delay elements of the uniform filter-bank are replaced by allpass filters of first order

$$z^{-1} \rightarrow A(z) = \frac{1-\alpha\, z}{z-\alpha} \quad \text{with} \quad \alpha \in \left\{\mathbb{R} \big| \, |\alpha| < 1\right\} . \qquad (1)$$

This bilinear transformation causes a *frequency warping*, which can achieve a very good approximation of the Bark or equivalent rectangular bandwidth (ERB) frequency scale as shown in [15]. These frequency scales model the non-uniform frequency resolution of the human auditory system [16]. The relation between warping coefficient α and sampling frequency f_s to approximate the Bark scale is given by [15]

$$\widehat{\alpha} = 1.0674 \sqrt{\frac{2}{\pi} \arctan\left(0.05683 \, \frac{f_\mathrm{s}}{\mathrm{kHz}}\right)} - 0.1916. \qquad (2)$$

This equation results an allpass coefficient of $\alpha = 0.776$ for the considered sampling frequency of $f_s = 48\,\mathrm{kHz}$. An approximation of the ERB scale can be achieved in a similar manner and the *uniform* filter-bank is simply obtained for $\alpha = 0$ according to Eq. (1).

The allpass transformation of Eq. (1) can be applied to a DFT filter-bank as well as a DCT filter-bank [13, 14]. The DFT analysis filters are obtained by a complex modulation of a prototype filter

$$h_i(n) = p_0(n)\, \exp\left\{ j\, \frac{2\pi}{M_{\mathrm{dft}}}\, i\, n \right\} \tag{3}$$

with sample index $n = 0, 1, \ldots, L-1$ and subband index $i = 0, 1, \ldots, M_{\mathrm{dft}} - 1$. The used FIR prototype filter $p_0(n)$ is an M-th band filter of length $L = 2\, M_{\mathrm{dft}}$ given by

$$p_0(n) = \frac{\mathrm{win}(n)}{M_{\mathrm{dft}}}\, \mathrm{si}\left(\frac{\pi}{M_{\mathrm{dft}}} \left(n - \frac{L}{2} \right) \right) \tag{4}$$

with $\mathrm{si}(n) = \sin(n)/n$ and $\mathrm{win}(n)$ marking the (Hann) window of length L.

The analysis filters of the considered (type-IV) DCT filter-bank are given by

$$h_i(n) = 2\, p_0(n) \cos\left(\frac{\pi}{M_{\mathrm{dct}}} (i + 0.5) \left(n - \frac{L-1}{2} \right) + (-1)^i\, \frac{\pi}{4} \right) \tag{5}$$

with subband index $i = 0, 1, \ldots, M_{\mathrm{dct}} - 1$. The FIR prototype filter $p_0(n)$ of length $L = 2\, M_{\mathrm{dct}}$ is designed by the approach of [17]. The DFT and DCT filter-bank can be both efficiently implemented by means of a polyphase network, e.g., [7].

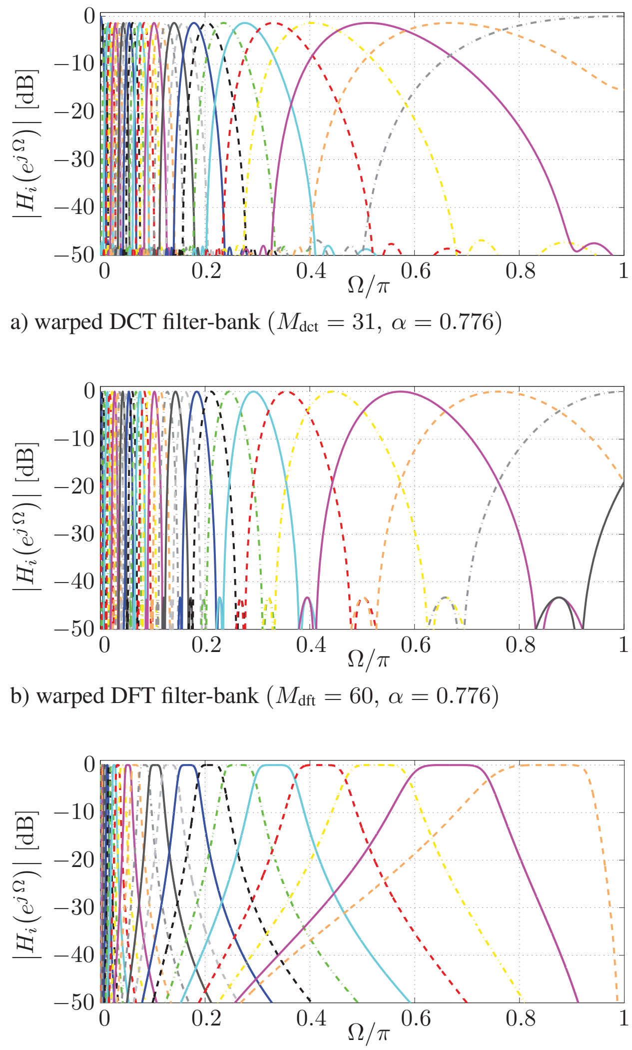
For the 1/3-octave filter-bank, Butterworth filters of 6-th order are used which fulfill the design specifications for class 1 filters according to [6].

Fig. 1 shows the magnitude responses of the analysis filters for the considered allpass transformed DFT and DCT filter-bank and the 1/3-octave filter-bank. The DCT filter-bank has $M_{\mathrm{dct}} = 31$ channels, the DFT filter-bank $M_{\mathrm{dft}} = 60$ channels and the 1/3-octave filter-bank possesses $M_{\mathrm{oct}} = 30$ channels. The three filter-banks are designed such that they have all the same number of unique frequency bands.[1] Fig. 1 reveals that the frequency bands of the warped[2] filter-banks are smaller for higher frequencies in comparison to those of the 1/3-octave filter-bank and vice versa for the lower frequency bands.

3. FREQUENCY DEPENDENT RT ESTIMATION

The RT can be determined in the time-domain from a measured RIR $h_{\mathrm{R}}(t)$ by means of the Schroeder integral [5]. The logarithm of the *energy decay curve* (EDC)

$$I_{\mathrm{S}}(t) = 10\, \log_{10} \int_t^{\infty} h_{\mathrm{R}}^2(\tau)\ \mathrm{d}\,\tau \tag{6}$$



a) warped DCT filter-bank ($M_{\mathrm{dct}} = 31$, $\alpha = 0.776$)

b) warped DFT filter-bank ($M_{\mathrm{dft}} = 60$, $\alpha = 0.776$)

c) 1/3-octave filter-bank ($M_{\mathrm{oct}} = 30$)

Fig. 1. Magnitude responses of different analysis filters with $\Omega = 2\pi\, f/f_s$.

is approximated by a linear function

$$f_{\mathrm{l}}(t) = b\, t + c \quad \text{for} \quad t_0 \le t \le t_1 \tag{7}$$

such that the (estimated) RT is given by $T_{60} = 60/b\,[\mathrm{s}]$. The parameters b and c are determined by a least-squares (LS) fit using, e.g., the MATLAB function `polyfit`. The time interval $[t_0, t_1]$ corresponds to the interval where the normalized EDC $\bar{I}_{\mathrm{S}}(t) = I_{\mathrm{S}}(t) - I_{\mathrm{S}}(0)$ declines from $-5\,\mathrm{dB}$ to $-35\,\mathrm{dB}$. The *normalized least-squares error* (NLSE)

$$\epsilon = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \left(I_{\mathrm{S}}(\tau) - f_{\mathrm{l}}(\tau) \right)^2\ \mathrm{d}\,\tau \tag{8}$$

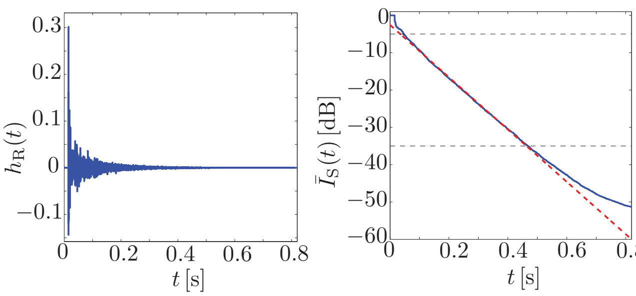
is used here as *reliability measure* for the estimated RT value.

The outlined RT estimation is exemplified in Fig. 2. The RIR is taken from the AIR database [18].[3] It has been measured in a

[1] A DFT of even size M provides complex subband signals, but has only $M_{\mathrm{dft}}/2 + 1$ unique subbands bands for a real input sequence. Therefore, a DCT filter-bank with $M_{\mathrm{dct}} = M_{\mathrm{dft}}/2 + 1$ subbands is used. The number of DCT bands M_{dct} is not identical to the number of octave filters M_{oct} as the octave filter-bank does not cover the region at $\Omega = \pi$ in comparison to the DCT filter-bank (see Fig. 1).

[2] The terms *frequency warping* and *allpass transformation* are used interchangeably as only allpass transformed filter-banks are considered.

[3] available at `http://www.ind.rwth-aachen.de/AIR`



a) RIR b) EDC with regression line

Fig. 2. RT estimation from a measured RIR. The blue solid line of subplot b) shows the normalized energy decay curve (EDC) and the red dotted line the regression line of Eq. (7).

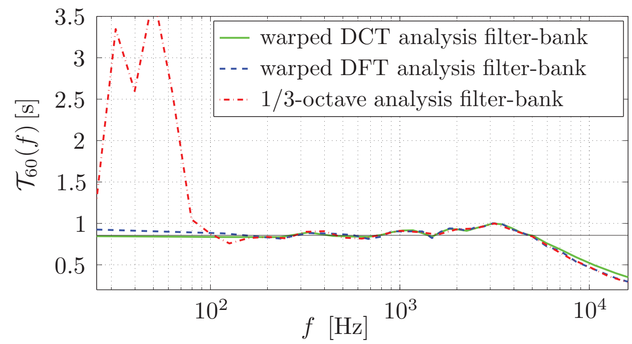
lecture room with a source-receiver distance of 5.56 m and without a dummy head ($f_s = 48$ kHz). A RT of $T_{60} = 0.86$ sec is calculated with a NLSE of $\epsilon = 6.96 \cdot 10^{-2}$.

For the estimation of the *frequency dependent* RT $\mathcal{T}_{60}(f)$, the described method is applied to a RIR after being filtered by the analysis filter-banks described in Sec. 2. For the DFT filter-bank with complex subband signals, the squared magnitude of the spectral coefficients is taken to calculate the EDC. The used 1/3-octave filter-bank performs a zero-phase filtering (by means of the MATLAB function `filtfilt`). This reduces the estimation error at lower frequencies [10], but causes also an increased computational load (and signal delay).

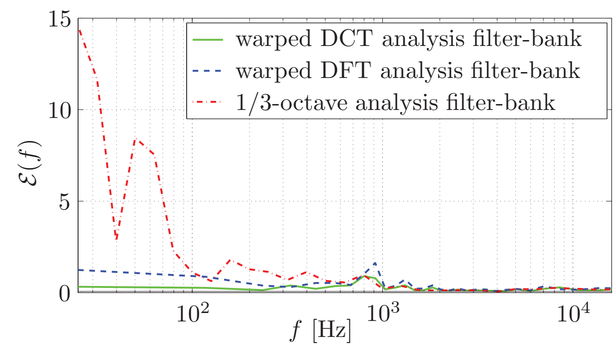
For the RIR of Fig. 2, the frequency dependent RT $\mathcal{T}_{60}(f)$ and NLSE for each frequency band $\mathcal{E}(f)$ are plotted in Fig. 3.[4] It can be seen that the 1/3-octave filter-bank provides much less reliable RT values at lower frequencies than the warped filter-banks. This is also reflected by the average NLSE value over all frequency bands, which amounts to 1.997 for the octave filter-bank where the values for the warped DCT and DFT filter-bank are equal to 0.281 and 0.596, respectively.

The above experiment has been conducted for 18 different RIRs of the AIR database. The RIRs are measured at different source-receiver distances and within different rooms (studio booth, office room, meeting room, stairway hall, corridor). The RTs are within the range $0.2\,\text{s} \leq T_{60} \leq 1.6\,\text{s}$. The frequency dependent NLSE averaged over all 18 measurements $\mathcal{E}_{\text{av}}(f)$ is plotted in Fig. 4. The 1/3-octave filter-bank exhibits again a significantly higher LS error at lower frequencies than the warped filter-banks. The error for the warped DCT filter-bank in turn is lower than for the warped DFT filter-bank. Averaging the values of $\mathcal{E}_{\text{av}}(f)$ over all frequency bands yields an error value of 0.857 for the DCT filter-bank, a value of 1.097 for the DFT filter-bank and a value of 2.402 for the 1/3-octave filter-bank. These different error values can be explained by the fact that very small filter bandwidths cause a high estimation error, cf., [8]. The 1/3-octave filters have very narrow bandwidths at low frequencies, which reasons the high error. The error value for the DCT filter-bank is the lowest one since its bandwidths are higher as for the DFT filter-bank (see Fig. 1).

[4] Even though a digital processing is performed, time-domain sequences are plotted over time t (as in Fig. 2) and frequency-domain quantities over frequency f (as in Fig. 3) to ease the physical interpretation.



a) frequency dependent RT



b) normalized LS error

Fig. 3. Calculation of the frequency dependent RT by means of different analysis filter-banks. The black solid line marks the RT of $T_{60} = 0.86$ s in the upper subplot and the NLSE of $\epsilon = 6.96 \cdot 10^{-2}$ in the lower subplot.

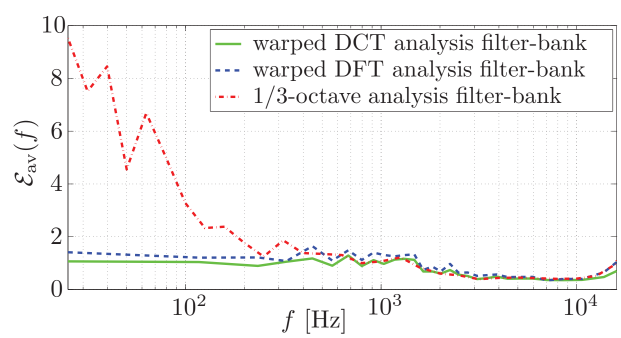


Fig. 4. Average normalized LS error $\mathcal{E}_{\text{av}}(f)$ for the estimation of the frequency dependent RT from 18 different RIRs.

4. ML-BASED RT ESTIMATION

A distinctive difference of the DCT filter-bank in comparison to the DFT filter-bank is that it decomposes a real input signal into *real* subband signals. This allows to apply algorithms in the frequency-domain which have been developed for the RT estimation in the time-domain. An important example is the RT calculation by means of a *maximum-likelihood* (ML) estimation [19–21]. These ML based techniques allow either to calculate the RT from a measured RIR (non-blind RT estimation) or to calculate the RT from a rever-

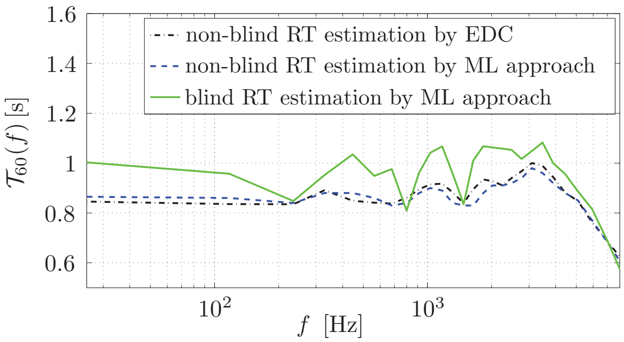


Fig. 5. Blind and non-blind estimation of the frequency dependent RT using a warped DCT filter-bank ($f_s = 16$ kHz).

berant speech signal (blind RT estimation). A blind RT estimation is of importance, if a dedicated measurement setup cannot be used as, for example, in the case of speech dereverberation systems, cf., [2, 3].

An example for these different RT estimations is shown in Fig. 5. The RIR of Fig. 2 downsampled to $f_s = 16$ kHz is considered and the frequency dependent RT is determined by means of the warped DCT filter-bank. The non-blind RT estimations by the EDC and the ML estimation of [20] yield similar curves for the frequency dependent RT $\mathcal{T}_{60}(f)$. For the blind RT estimation, a clean speech signal of 15 s duration and 16 kHz sampling frequency is convolved with the considered RIR. The reverberant speech is filtered by the warped DCT filter-bank and the blind RT estimation of [21] is applied in each subband. The solid curve in Fig. 5 shows the average RT value for each frequency band (as the algorithm in [21] is designed to track time-varying RTs). The comparison with the non-blind estimation reveals an accuracy of about 150 ms, which is similar to the accuracy that is achieved for a blind RT estimation in the time-domain, cf., [20, 21].

It is also conceivable to estimate the frequency dependent RT in noisy environments by applying the approach of [20] in the frequency-domain. However, an elaboration of this case exceeds the scope of this work.

5. CONCLUSIONS

An improved approach for the estimation of the frequency dependent RT by means of an allpass transformed DCT or DFT filter-bank is presented. These warped filter-banks achieve a much better approximation of the non-uniform frequency resolution of the human ear than commonly used octave filter-banks. Moreover, a uniform or non-uniform (Bark or ERB) frequency resolution is simply adjusted by a single allpass coefficient and the warped DCT or DFT filter-bank can be efficiently implemented by means of a polyphase network. It is shown that warped filter-banks estimate the frequency dependent RT with a much lower error than commonly used 1/3-octave filter-banks. The warped DCT filter-bank is of special interest as it provides real-valued subband signals. This allows to apply an ML estimator for either a non-blind estimation of the frequency dependent RT from a RIR or a blind RT estimation from a reverberant speech signal.

6. REFERENCES

[1] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, UK, 4th edition, 2000.

[2] K. Lebart, J. M. Boucher, and P. N. Denbigh, “A New Method Based on Spectral Subtraction for Speech Dereverberation,” *acta acoustica - ACOUSTICA*, vol. 87, no. 3, pp. 359–366, 2001.

[3] E. A. P. Habets, *Single- and Multi-Microphone Speech Dereverberation using Spectral Enhancement*, Ph.D. thesis, Eindhoven University, Eindhoven, The Netherlands, 2007.

[4] ISO-3382, “Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters,” International Organization for Standardization, Geneva, Switzerland, 1997.

[5] M. R. Schroeder, “New Method of Measuring Reverberation Time,” *Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[6] American National Standards Institute, “*Specification for Octave-Band and Fraction-Octave-Band Analog and Digital Filters*,” ANSI S1.11-1986, 1986.

[7] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Upper Saddle River, New Jersey, 1993.

[8] F. Jacobsen, “A Note on Acoustic Decay Measurement,” *Journal of Sound and Vibration*, vol. 115, no. 1, pp. 163–170, 1987.

[9] S.-K. Lee, “An Acoustic Decay Measurement Based on Time-Frequency Analysis Using Wavelet Transform,” *Journal of Sound and Vibration*, vol. 252, no. 1, pp. 141–153, 2002.

[10] F. Jacobsen and J. H. Rindel, “Time Reversed Decay Measurement,” *Journal of Sound and Vibration*, vol. 117, no. 1, pp. 187–190, 1987.

[11] J.-M. Jot, “An Analysis/Synthesis Approach to Real-Time Artificial Reverberation,” in *Proc. of Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, San Francisco (California), USA, Mar. 1992, pp. 221–224.

[12] C. Braccini and A. V. Oppenheim, “Unequal Bandwidth Spectral Analysis using Digital Frequency Warping,” *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 22, no. 4, pp. 236–244, Aug. 1974.

[13] G. Doblinger, “An Efficient Algorithm for Uniform and Nonuniform Digital Filter Banks,” in *Proc. of Intl. Symposium on Circuits and Systems (ISCAS)*, Singapore, June 1991, vol. 1, pp. 646–649.

[14] N. I. Cho and S. K. Mitra, “Warped Discrete Cosine Transform and Its Application in Image Compression,” *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 8, pp. 1364–1373, Dec. 2000.

[15] J. O. Smith and J. S. Abel, “Bark and ERB Bilinear Transforms,” *IEEE Trans. on Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, Nov. 1999.

[16] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, London, UK, 4th edition, 1997.

[17] Y.-P. Lin and P. P. Vaidyanathan, “A Kaiser Window Approach for the Design of Prototype Filters for Cosine Modulated Filterbanks,” *IEEE Signal Processing Letters*, vol. 5, no. 6, pp. 132–134, June 1998.

[18] M. Jeub, M. Schäfer, and P. Vary, “A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms,” in *Proc of. Intl. Conference on Digital Signal Processing (DSP)*, Santorini, Greece, July 2009.

[19] R. Ratnam, D. L. Jones, B. C. Wheeler, W. D. O’Brien, C. R. Lansing, and A. S. Feng, “Blind Estimation of Reverberation Time,” *Journal of the Acoustical Society of America*, vol. 114, no. 5, pp. 2877–2892, Nov. 2003.

[20] H. W. Löllmann and P. Vary, “Estimation of the Reverberation Time in Noisy Environments,” in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle (Washington), USA, Sept. 2008.

[21] H. W. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, “An Improved Algorithm for Blind Reverberation Time Estimation,” in *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.