BLIND ESTIMATION OF REVERBERATION TIME BASED ON SPECTRO-TEMPORAL MODULATION FILTERING

Feifei Xiong†*, Stefan Goetze*† *and Bernd T. Meyer*‡

†Fraunhofer Institute for Digital Media Technology (IDMT), Project Group
Hearing-, Speech- and Audio-Technology (HSA), 26129 Oldenburg, Germany
Email: {feifei.xiong,s.goetze}@idmt.fraunhofer.de
‡Medical Physics, Institute of Physics, University of Oldenburg, 26111 Oldenburg, Germany
Email: bernd.meyer@uni-oldenburg.de

ABSTRACT

A novel method for blind estimation of the reverberation time (RT60) is proposed based on applying spectro-temporal modulation filters to time-frequency representations. 2D-Gabor filters arranged in a filterbank enable an analysis of the properties of temporal, spectral, and spectro-temporal filtering for this task. Features are used as input to a multi-layer perceptron (MLP) classifier combined with a simple decision rule that attributes a specific RT60 to a given utterance and allows to assess the reliability of the approach for different resolutions of RT60 classification. While the filter set including temporal, spectral, and spectro-temporal filters already outperforms an MFCC baseline, the error rates are further reduced when relying on diagonal spectro-temporal filters alone. The average error rate is 1.9% for the best feature set, which corresponds to a relative reduction of 58.3% compared to the MFCC baseline for RT60s in 0.1 s resolution.

Index Terms— Blind reverberation time estimation, spectro-temporal modulation, 2D Gabor filterbank

1. INTRODUCTION

The effect of reverberation significantly decreases the perceived speech quality [1] and speech recognition accuracy [2], especially in distant microphone scenarios. This is caused by multi-path propagation of an acoustic signal from its source to the microphone, usually represented by the room impulse response (RIR). Reverberation time (RT60) is an important characteristic of the RIR; it is defined as the time interval for a 60 dB decay of the sound energy after switching off the sound source and is commonly used as a necessary input for dereverberation techniques in the field of speech enhancement [3, 4] and for automatic speech recognition (ASR) systems [5, 6]. The reverberation time strongly correlates with the amount of *perceived* reverberation [7].

Different approaches for RT60 estimation have been proposed in the literature. Schroeder [8] introduced the backward integration method to estimate RT60 directly from a measured RIR in 1965. A method to obtain RT60 from a sound decay by using interrupted noise is described in [9]. Recently several algorithms for *blind* RT60 estimation have been developed, which estimate RT60 from the observed reverberant speech signal without knowledge of the room geometry or a reference signal. Based on the spectral decay distribution of the reverberant signal, RT60 is predicted in [10] by estimating the decay rate at each frequency band in the discrete Fourier transform (DFT) domain. A comparison of energies at high and low modulation frequencies, the so-called speech-to-reverberation modulation energy ratio (SRMR), which is highly correlated to RT60 is evaluated in [11]. In [12] a blind RT60 estimation is achieved by a statistical model of the sound decay characteristics of reverberant speech. This work was extended by [13] using a pre-selection mechanism to detect plausible decays and a subsequent application of a maximum-likelihood (ML) criterion to estimate RT60 in the time domain.

Motivated by the progress that has been achieved using artificial neural networks in machine learning tasks, [14] proposed another method to estimate RT60 blindly from trained reverberant speech, for which short-term root-mean square (RMS) values of speech signals were used as the neural network input. In [15] an adaptive hidden Markov model (HMM) was used to determine RT60 within an iterative search of the HMM sequence ML by decreased or increased RT60 values based on an initial RT60 estimate. An RT60 classifier based on Gaussian mixture models (GMMs) using mel-frequency cepstral coefficient (MFCC) features [16] as input has been introduced by [6] to estimate the late reverberation energy for ASR systems. However, these studies are either restricted by the energy envelope of the word-level speech utterance that is easily corrupted by e.g. speech tone bursts or fluctuating noise, or do not report the actual performance of RT60 estimation, e.g. in terms of the error rate. In this study, an alternative approach for blind RT60 estimation is presented and evaluated on the basis of spectro-temporal modulation filtering in the short-time Fourier transform (STFT) domain. Since reverberation results in a modulation spectrum that is shifted towards lower modulation frequencies [17], we expect these changes to be reflected in the output of a spectro-temporal filterbank. The output of this filterbank is used as input to an artificial neural network with one hidden layer (multi-layer perceptron, MLP), which is trained for blind estimation of RT60s. The spectro-temporal modulation features are obtained using a two-dimensional (2D) Gabor filterbank that has been shown to result in robust features that have successfully been applied to speech classification [18, 19, 20].

The remainder of this paper is organized as follows: Section 2 introduces the 2D Gabor filterbank and describes the effect of reverberation on the filterbank output. The MLP classifier and the decision rule that converts the MLP-output from frame-wise to utterance-

This work was partially supported by the project MOBECS (AAL-2011-4-038) co-funded by the European Commission (EC) and the Federal Ministry of Education and Research (BMBF) in the Ambient Assisted Living (AAL) program and the EU-FP7 project S4EeB (grant agreement no. 284628), as well as by the DFG (SFB/TRR 31 'The active auditory system'; URL: http://www.uni-oldenburg.de/sfbtrr31).

wise are briefly described in Section 3. The experimental procedure and results are presented in Section 4 for the proposed RT60 estimation. The discussion and summary in Section 5 conclude the paper.

2. 2D GABOR FILTERBANK

Physiological measurements in the primary auditory cortex of different mammal species showed that neurons are sensitive to specific spectro-temporal patterns (e.g. the neurons exhibit high firing rates only when a specific modulation frequency is represented in the stimulus), which are referred to as spectro-temporal receptive fields (STRFs). 2D Gabor filters have been successfully applied for modeling STRFs [21], which motivated their use as front-end for auditory-inspired robust ASR [18, 20]. This is achieved by filtering a time-frequency representation (in our case a log-mel-spectogram) with several filters, with the aim of explicitly capturing spectro-temporal modulations (e.g. diagonal patterns in the spectrogram).

In the present study, we rely on the filterbank outlined in [20] that is based on Gabor filters, represented as follows,

$$\mathbf{G}[k,n] = \mathbf{s}_{\mathrm{carr}}[k,n] \cdot \mathbf{h}_{\mathrm{env}}[k,n], \quad (1)$$

$$\mathbf{s}_{\mathrm{carr}}[k,n] = \exp\left[i\omega_k(k-k_0) + i\omega_n(n-n_0)\right], \quad (2)$$

$$\mathbf{h}_{\mathrm{env}}[k,n] = 0.5 - 0.5 \cdot \cos\left[\frac{2\pi(k-k_0)}{W_k+1}\right] \cdot \cos\left[\frac{2\pi(n-n_0)}{W_n+1}\right], \quad (3)$$

where k and n denote the spectral and temporal indices, and W_k, W_n are the envelope window lengths with the center indices k_0, n_0, respectively. The periodicity of the carrier function is defined by the radian frequencies ω_k, ω_n, which allow the Gabor filters to be tuned to particular directions of spectro-temporal modulation. Purely spectral and temporal modulation filtering can be performed by choosing $\omega_k = 0$ or $\omega_n = 0$, respectively.

The temporal and spectral modulation frequencies have been found to be relevant in speech and signal processing applications, which cover temporal modulations from 2 to 25 Hz and spectral modulations from -0.25 to 0.25 cycle/channel, respectively. The spacing of the filters is chosen so that all filters evenly cover the modulation frequency domain which results in an unbiased representation of the frequencies. The resulting filters as used in [22] are shown in Fig. 1. Filtering the log-mel-spectrogram (23 frequency channels) with all 59 filters results in a very high feature dimensionality ($23 \times 59 = 1357$), which is reduced to 449 dimensions by performing a subsampling of frequency channels. For filters with a large spectral extent, highly-correlated values are obtained for neighboring channels of the filter output. A reduction is achieved by selecting only the center frequency channel and the channels with a distance of $k \cdot \frac{S}{4}$ where S is the spectral extent of the filter. This is performed for each Gabor filter, and the individual filter outputs are stacked to form the final feature vector.

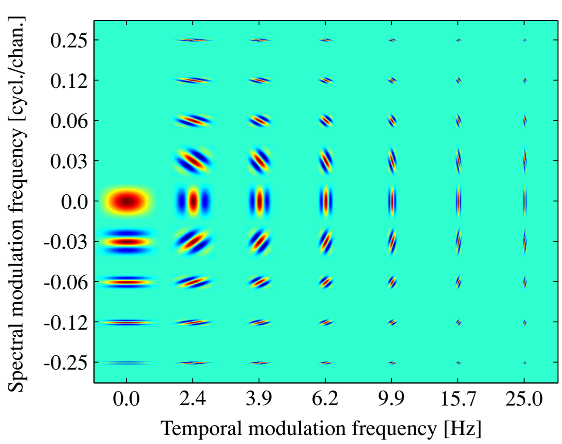

Fig. 1. Real component of the Gabor filterbank set [22], which contains 59 Gabor filters for different spectral, temporal and spectro-temporal modulation frequencies.

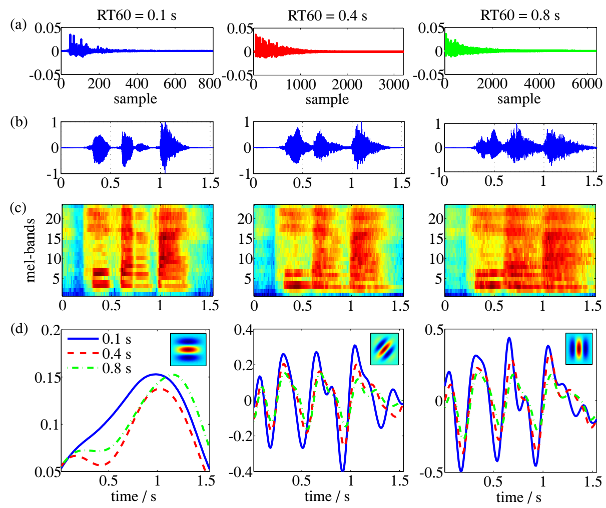

Fig. 2. (a) RIRs with RT60s of 0.1, 0.4 and 0.8 s at 8 kHz sampling rate; (b) reverberant signal from clean digit utterance "two, seven, five" with RIRs w.r.t. (a); (c) the corresponding log-mel-spectrogram of (b); (d) the output feature coefficients by one purely spectral (left), one spectro-temporal (mid) and one purely temporal (right) Gabor filter from Fig. 1 (-0.03 cycle/channel and 3.9 Hz).

Fig. 2 exemplary shows how the output of the Gabor filterbank is affected by RIRs with different RT60s. It can be seen from Fig. 2 (c) that reverberation smears the spectrogram in both spectral and temporal modulation domains. Row (d) shows examples of the filter output after processing with a purely spectral (left), a spectro-temporal (mid) and a purely temporal (right) filter, respectively. All of them are affected by the reverberation caused by the RIRs having different RT60s shown in row (a). In general, shifting and compression effects on the feature values by reverberation can be observed. It seems that the *shifting* plays a dominant role for spectral modulation filtering as RT60 increases, while the *compression* of energy seems to be the dominating factor for temporal modulations. Both effects can be observed for the spectro-temporal modulation filter. These hints at the discriminative power of the filters incorporated in the filterbank for RT60 estimation are analyzed in the following sections.

3. SPEECH DATA AND CLASSIFIER

Multi-layer perceptrons (MLPs) belong to the class of feedforward artificial neural network models and serve as classifier converting the input features to posteriori probabilities for different classes of RT60 values. The final (frame-wise) classification result is obtained with a simple *winner-take-all* decision rule that is attributed to the whole input utterance.

The MLP is trained with the backpropagation algorithm [23]. The number of neurons in the input layer is 449 (Gabor feature components) $\times$ 9 (frames of temporal context). The number of hidden units is 160, while the number of output units depends on the amount of RT60 classes to estimate. The parameters for temporal context and the number of hidden units were the same as for the ASR experiments in [22], i.e. they were not optimized for a specific feature set used in this study. MLP training is carried out with a reverberated version of Aurora 2 [24] with frame-wise labels for the RT60 corresponding to that utterance. After the training procedure, frame-wise posterior probabilities for each class are obtained by the MLP forward run, where frame-wise probabilities for each RT60 class are merged to a single decision by temporal application and application of a winner-take-all decision rule, as shown in Fig. 3 for instance.

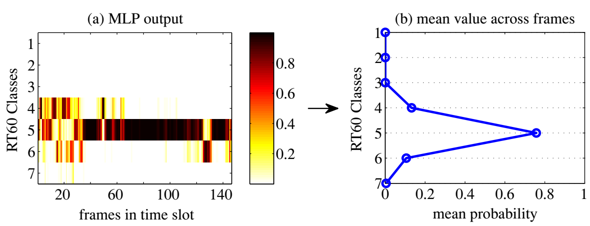

Fig. 3. Visualization of the decision rule applied to convert the frame-wise MLP-output (left, color denotes class probability) to a RT60 class for each utterance. The mean value of the probabilities across time frames is calculated for each RT60 class, and the class with the maximum class probability is chosen. For the seven-class problem shown here, possible RT60 values range from 0.05 to 3.2 s.

4. EXPERIMENTS AND RESULTS

In order to generate the reverberant speech database with various RT60s for MLP training, the noise-free part of the Aurora 2 corpus [24] is chosen as the prototype to convolve with different RIRs. It contains 8440 digit utterances for training and 4004 for testing at sampling rate of 8 kHz. The RIR generator for different RT60s is implemented based on the image method proposed in [25] (3 examples being shown in Fig. 2 (a)). Overlapping speech segments of 25 ms with 10 ms shift are applied for feature extraction by the real part of 2D Gabor filters, where the spectro-temporal modulation filtering is separated into 5 groups from Fig. 1, i.e. (i) purely spectral modulation by 5 Gabor filters (dimension of 35), (ii) purely temporal modulation by 7 Gabor filters (dimension of 7), (iii) combination of (i) and (ii) (dimension of 41), (iv) spectro-temporal modulation *excluding* (i) and (ii) (dimension of 408), and (v) spectro-temporal modulation *including* (i) and (ii) (dimension of 449). As a comparison, the baseline MFCC [16] with delta and double-delta coefficients (dimension of 39) is employed in our experiments. MLP implementation is provided by the QuickNet software package (c.f. http://www1.icsi.berkeley.edu/Speech/qn.html).

Fig. 4 gives an overview of the experimental setup for RT60 estimation in MLP network. The reverberant database is generated by convolution of the clean utterances and RIRs with different RT60s ($x = s * h$), which are then converted to features f and used to train and forward-run the MLP. The frame-wise probabilities p for the RT60 classes are subject to the decision rule, effectively producing a decision for the utterance, as illustrated in Fig. 3.

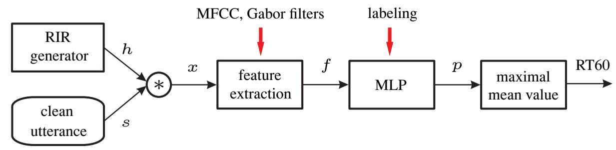

Fig. 4. Overview of the experimental setup for RT60 estimation.

4.1. RT60 Estimation Accuracy

To evaluate the RT60 estimation accuracy based on spectro-temporal processing, RIRs with 7 different RT60s are used in this experiment, i.e. {0.05, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2} s, resulting in 7 classes for MLP labeling. We present two measures to assess the accuracy: the absolute error rate (ER), as well as the maximal mean probability that reflects the certainty of the classifier. Fig. 5 shows estimation ERs below 0.5% and maximal mean probabilities around 0.9 for purely spectro-temporal (iv) and the complete filter set (v), reducing the baseline ERs by more than 56% relative on average. Compared all other feature sets, purely spectral (i) and temporal (ii) filtering alone strongly increase ERs. Even though the purely spectral and purely temporal filtering are combined together, the results for set (iii) are not comparable to the sets including spectro-temporal filtering. However, the results show a complementarity of spectral and temporal modulations for the analyzed range of RT60s. The increasing ER with ascending RT60 values for set (i) indicates that pure spectral features are not sensitive to longer RT60s, e.g. >0.4 s. On the other hand, purely temporal features (ii) become more discriminative with *longer* RT60s. We assume that reverberation for

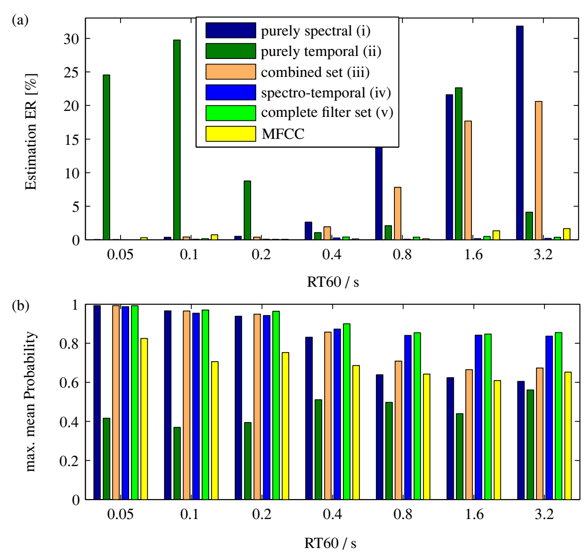

Fig. 5. (a) RT60 estimation ER for the trained RIRs from 0.05 to 3.2 s increasing at a factor of two for different feature extraction methods; (b) the corresponding maximal mean probabilities from MLP output for the class decision.

RT60s <0.4 s has a predominant effect on spectral modulations, while RT60s with higher values have a stronger effect on temporal modulations. This may also indicate that shifting effect on the spectro-temporal features should be paid more attention for small reverberation with short RT60s, and compression effect needs more analysis for large reverberation with long RT60s, corresponding to Fig. 2 (d).

feature methods	error rate	rel. imp.	max. mean prob.	rel. imp.
MFCC	0.64	0.0	0.70	0.0
purely spectral (i)	10.37	-1520	0.80	14.3
purely temporal (ii)	13.28	-1975	0.46	-34.3
combined set (iii)	6.99	-992	0.83	18.6
spectro-temporal (iv)	**0.14**	**78.1**	0.90	28.6
complete filter set (v)	0.28	56.3	**0.91**	**30.0**

Table 1. Average ERs [%] and maximal mean probabilities, as well as the respective relative improvements [%] according to Fig. 5.

Table 1 presents the estimation ERs averaged over the RT60 classes and the corresponding maximal mean probabilities, as well as the relative improvements compared to the MFCC baseline. The results show that combination of the purely spectral and purely temporal filters improves the estimation performance slightly, compared to the purely spectral or temporal filtering alone. The filter set with spectro-temporal filters *only* performs best. The complete filter set is still better than the baseline system, but the inclusion of purely temporal and spectral filters (iv) somewhat increases ERs. Although MFCCs perform a spectral modulation filtering and also a temporal filtering (due to the inclusion of deltas and double-deltas), which can be described as a (separable) spectro-temporal filter, it seems that the explicit spectro-temporal processing capturing diagonal structures is an important aspect for this task, highlighted by the relative improvement of 78.1% over the MFCC baseline.

4.2. Resolution of RT60 Estimation

Depending on the application using the RT60 estimate, the required resolution for the RT60 estimate differs. In this experiment we evaluate the properties of the proposed algorithm based on the aforementioned results considering high RT60 resolutions. We chose the best performing filter sets (iv) and (v), as well as the MFCC baseline for this assessment. Two higher resolution sets of RT60s for equally spaced RIRs were generated using the method described above with RT60s ranging from 0.1 to 1.0 s at a 0.1 s-stepsize, with the corresponding 10-class labels for MLP-training and testing, respectively. Please note that direct-to-reverberant energy ratio (DRR) was kept the same for RIRs with identical RT60.

feature methods	error rate	rel. imp.	max. mean prob.	rel. imp.
MFCC	4.56	0.0	0.46	0.0
spectro-temporal (iv)	**1.90**	**58.3**	0.68	47.8
complete filter set (v)	3.81	16.5	**0.69**	**50.0**

Table 2. Average ERs [%] and maximal mean probabilities, as well as the relative improvements [%] according to Fig. 6 (where maximal mean probability plot is omitted due the same trend as Fig. 5 (b)).

As shown in Fig. 6 (a), results are consistent with ERs reported in Section 4.1. The spectro-temporal set (iv) still performs best,

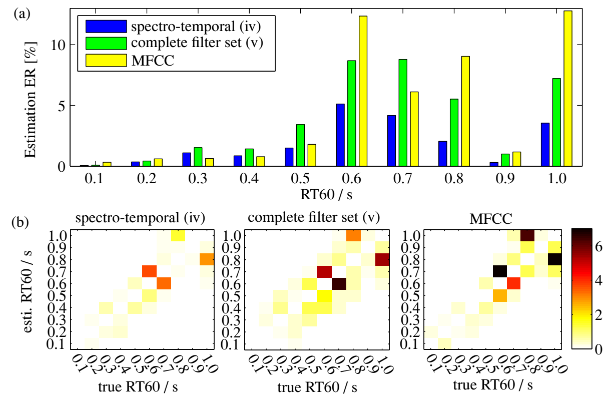

Fig. 6. RT60 estimation performance for the trained RT60s from 0.1 to 1.0 s with stepsize of 0.1 s. (a) estimation ER with spectro-temporal (iv), complete filter set (v) and MFCC; (b) error distributions of estimated RT60s versus true RT60s (Diagonal elements have been set to 0 for visibility of the error rates).

with an average ER of 1.9% (cf. Table 2), and inclusion of temporal/spectral filters (v) results in ERs between filter set (iv) and the MFCC baseline. The confusion matrices in Fig. 6 (b) show that a very small RT60 estimation variance can be achieved by the spectro-temporal set (iv). It is also of interest to notice that the performance of RT60 estimation in Table 2 decreases compared to the previous experiment results in Table 1. This can be explained by the fact that discrimination amongst the spectro-temporal features is reduced by the more similar reverberation effects in a higher RT60 resolution.

5. DISCUSSION AND SUMMARY

This contribution presented an alternative method for blind RT60 estimation by applying spectro-temporal modulation filtering using a multi-layer perceptron (MLP) as classifier. Spectro-temporal modulation filtering has been applied as a robust estimator for RT60s from a reverberant speech signal. An average error rate of 1.9% (compared to an error rate of 3% but within 0.1 s estimation tolerance mentioned in [14]) indicates a significant improvement achieved by the proposed RT60 estimation method. Furthermore, the computational complexity is reduced compared to [14] in which an additional second decision-based watchdog neural network was used. Evaluations for various RIRs with different RT60s address more detailed performance analysis than e.g. in [15, 6], that focus on the acoustic model adaptation using an approximated RT60 estimation in ASR systems. The proposed method is characterized by small variance of the estimations errors which makes the results comparable to state-of-the-art blind RT60 estimation approaches [10, 11, 13, 26] (of interest also refer to [27] with results of performance comparison amongst these state-of-the-art schemes).

Compared to the purely spectral and purely temporal modulation filter sets, such as MFCCs, spectro-temporal modulation filters are sensitive (and hence discriminative) in presence of reverberation over a wide range of RT60s. While purely spectral and temporal filters had a detrimental effect on the overall performance, it is interesting to note that these filters have complementary properties, i.e. are either sensitive to long or short RIRs. Due to the reliable RT60 estimation results the proposed method is suitable for dereverberation in speech enhancement and/or robust ASR systems.

6. REFERENCES

[1] S. Goetze, E. Albertin, J. Rennies, E.A.P. Habets, and K.-D. Kammeyer, "Speech Quality Assessment for Listening-Room Compensation," in *38th Audio Engineering Society (AES) Conf.*, Pitea, Sweden, Jul. 2010, pp. 11–20.

[2] A. Sehr, *Reverberation Modeling for Robust Distant-Talking Speech Recognition*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, Oct. 2009.

[3] K. Lebart, J. Boucher, and P.N. Denbigh, "A New Method based on Spectral Subtraction for Speech Dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, 2001.

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

[5] K. Kinoshita, T. Nakatani, and M. Miyoshi, "Efficient Blind Dereveration Framework for Automatic Speech Recognition," in *Proc. Interspeech*, 2005, pp. 3145–3148.

[6] R. Gomez and T. Kawahara, "Robust Speech Recognition based on Dereverberation Parameter Optimization using Acoustic Model Likelihood," *IEEE Trans. on Audio, Speech and Lang. Proc.*, vol. 18, no. 7, pp. 1708–1716, Sep. 2010.

[7] H. Kuttruff, *Room Acoustics*, Spon Press, London, 4th edition, 2000.

[8] M.R. Schroeder, "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Amer.*, vol. 37, no. 3, pp. 409–412, 1965.

[9] "ISO Norm 3382: Acoustics – Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters," *International Organization for Standardization (ISO), Geneva, Switzerland*, 1997.

[10] J.Y.C. Wen, E.A.P. Habets, and P.A. Naylor, "Blind Estimation of Reverberation Time based on the Distribution of Signal Decay Rates," in *Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, Apr. 2008, pp. 329–332.

[11] T. Falk, C-Zheng, and W.-Y. Chan, "A Non-intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 18, no. 7, pp. 1766–1774, 2010.

[12] R. Ratnam, D.L. Jones, B.C. Wheeler, J.W.D. O'Brien, C.R. Lansing, and A.S. Feng, "Blind Estimation of Reverberation Time," *J. Acoust. Soc. Amer.*, vol. 114, no. 5, pp. 2877–2892, 2003.

[13] H. Löllmann, E. Yilmaz, M. Jeub, and P. Vary, "An Improved Algorithm for Blind Reverberation Time Estimation," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aug. 2010.

[14] T.J. Cox, F. Li, and P. Dalington, "Extracting Room Reverberation Time from Speech Using Artificial Neural Networks," *J. Acoust. Soc. Amer.*, vol. 94, no. 4, pp. 219–230, 2001.

[15] H.G. Hirsch and H. Finster, "A New Approach for the Adaptation of HMMs to Reverberation and Background Noise," *Speech Commun.*, vol. 50, no. 3, pp. 244–263, 2008.

[16] S.B. David and P. Mermelstein, "Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. on Acoustics, Speech and Signal Proc.*, vol. 28, no. 4, pp. 357–366, 1980.

[17] T. Houtgast and H.J.M. Steeneken ., "A Review of the MTF Concept in Room Acoustics and Its Use for Estimating Speech Intelligibility in Auditoria," *J. Acoust. Soc. Amer.*, vol. 77, no. 3, pp. 1069–1077, 1985.

[18] M. Kleinschmidt and D. Gelbart, "Improving Word Accuracy with Gabor Feature Extraction," in *Interspeech*, 2002, pp. 25–28.

[19] V. Mildner, S. Goetze, K.-D. Kammeyer, and A. Mertins, "Optimization of Gabor Features for Text-Independent Speaker Identification," in *Int.Symp. on Circuits and Systems (ISCAS)*, New Orleans, USA, May 2007, pp. 3932–3935.

[20] B.T. Meyer and B. Kollmeier, "Robustness of Spectro-Temporal Features against Intrinsic and Extrinsic Variations in Automatic Speech Recognition," *Speech Commun.*, vol. 53, no. 5, pp. 753–767, 2011.

[21] A. Qiu, C.E. Schreiner, and M.A. Escabí, "Gabor Analysis of Auditory Midbrain Receptive Fields: Spectrotemporal and Binaural Composition," *J Neurophysiol*, vol. 90, no. 1, pp. 456–476, 2003.

[22] B.T. Meyer, S.V. Ravuri, M.R. Schädler, and N. Morgan, "Comparing Different Flavors of Spectro-Temporal Features for ASR," in *Interspeech*, 2011, pp. 1269–1272.

[23] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation," *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1: Foundations. MIT Press, 1986.

[24] D. Pearce and H. Hirsch, "The Aurora Experimental Framework for the Performance Evaluation of Speech Recognition Systems under Noisy Conditions," in *Int. Conf. on Spoken Lang. Proc.*, 2000, pp. 29–32.

[25] J.B. Allen and D.A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[26] N. López, Y. Grenier, G. Richard, and I. Bourmeyster, "Low Variance Blind Estimation of the Reverberation Time," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sep. 2012.

[27] N.D. Gaubitch, H. Löllmann, M. Jeub, T.H. Falk, P.A. Naylor, P. Vary, and M. Brookes, "Performance Comparison of Algorithms for Blind Reverberation Time Estimation from Speech," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Sep. 2012.