Model-based Optimization of a Low-dimensional Modulation Filter Bank for DRR and T60 Estimation

Semih Ağcaer
Institute of Communication Acoustics
Ruhr-Universität Bochum
Bochum, Germany
semih.agcaer@rub.de

Rainer Martin
Institute of Communication Acoustics
Ruhr-Universität Bochum
Bochum, Germany
rainer.martin@rub.de

***Abstract*—Amplitude Modulation Spectrum (AMS) features can be implemented as a cascade of two filter banks whereas the filter bandwidths can be optimized for a particular application. In this work we train AMS-based features using a combination of a model-based optimization (MBO) approach and feature selection for full-band DRR and full-band T_{60} estimation. MBO replaces the computational complex data-based cost function by approximating a less complex surrogate model and thus reduces the time needed for training. We evaluate our approach on the publicly available ACE challenge corpus and achieve with only five features the best RMSE in the DRR estimation task using the single microphone configuration and upper mid-range performance for T_{60} estimation. The computational complexity of our algorithm is much lower than all other submitted algorithms.**

***Index Terms*—DRR estimation, T_{60} estimation, amplitude modulation spectrum, model-based optimization**

I. Introduction

Accurate estimation of room acoustic parameters like the reverberation time (T_{60}) or the direct-to-reverberant ratio (DRR) are a challenging tasks. Information about T_{60} or DRR could be used in different signal processing algorithms, e.g. for dereverberation. The reverberation time of a room is defined as the time in which the acoustic energy density decreases by 60dB after switching the sound source off and depends in general on the room size and wall absorption coefficients. The DRR defines the energy ratio between the energy of the direct acoustic path to the energy of the remaining acoustic paths.

The ACE (acoustical characterization of environments) challenge was hosted with a view to compare different algorithms. The results indicate that state-of-the-art algorithms for T_{60} estimation mostly relying on decay rates can achieve the most accurate estimation results. Analytical approaches, e.g. by Prego et al. [1] and Löllmann et al. [2] outperformed other proposals in this task. However, DRR estimation performance in the ACE challenge was dominated by algorithms based on machine learning approaches. To estimate the DRR Parada et al. [3] trained a neural network with a set of different input features, e.g. line spectral frequencies (LSF), zero-crossing rate, parameters extracted from the power spectrum of long term deviation and modulation domain features.

Recent work of Xiong et al. [4] demonstrated the successful use of auditory-inspired acoustic features for room acoustic parameter estimation. Their approach is based on using temporal modulation features extracted from the time-frequency representation and on training a multi-layer perceptron (MLP) neural network for estimating the room acoustic parameters. In our previous work, we proposed a low-dimensional AMS-based feature extraction algorithm which we used for acoustic scene classification [5] and speaker localization [6]. The main idea is similar to the temporal modulation features used by Xiong et al. since they are also inspired by the human auditory signal processing and were successfully used for acoustic scene classification tasks before [7]–[10].

The proposed feature extraction bases mainly on two successive filter banks, non-linear operations and a final averaging step. Since both filter banks use only a small number of IIR filters the proposed approach is less complex than, e.g. [4]. We optimized the filter bandwidths of the two filter banks with a model-based optimization (MBO) method in order to estimate T_{60} and DRR. The number of filters we utilized to accomplish the two tasks differ. The T_{60} estimation performs better with a higher number of filters, whereas DRR estimation performs better with less filters.

In the next sections of this paper, we will introduce our proposed AMS-based feature extraction and the MBO training scheme. Then, in Section III we evaluate our proposed approach on the ACE challenge corpus and present the results. Section IV concludes this paper with a discussion of these results.

II. Methods

A. AMS Feature Extraction

Figure 1 depicts the block diagram of the proposed AMS feature extraction which differs slightly from the approach we used for the DCASE challenge 2013 [5]. It mainly consist of two successive filter banks. Before an input signal $x(k)$ with length N is fed into the first filter bank with N_{TF}

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

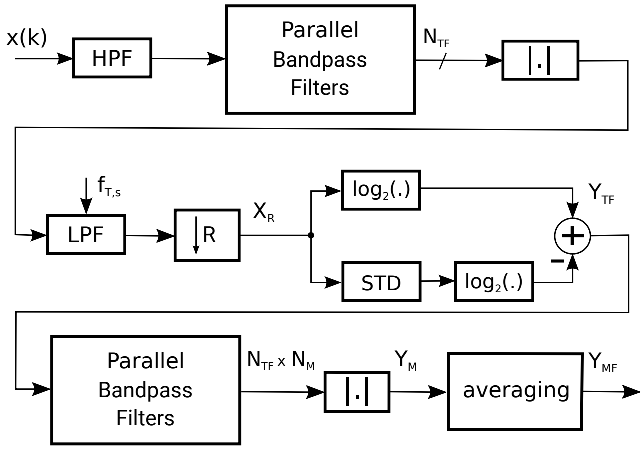

Fig. 1: Block diagram of AMS-based modulation-domain feature extraction.

parallel bandpass filters, it is highpass-filtered such that any DC components below $25\,\mathrm{Hz}$ are removed. The first of the N_{TF} filter is designed as a lowpass filter with the cutoff frequency f_{u1}. The following filters are designed as bandpass filters with lower edge frequencies equal to the upper edge frequencies of the preceding filter (see Figure 2). The output of this first filter bank are N_{TF} subband signals $x_{T,i}(k)$ with different spectral content. These subband signals are rectified and then lowpass filtered and decimated in the next two steps. The cutoff frequency $f_{T,s}$ of this lowpass filter is equal to the highest upper edge frequency of the second filter bank plus a margin of $40\,\mathrm{Hz}$. The decimation factor is $R = \lfloor \frac{f_s}{2 \cdot f_{T,s}} \rfloor$, where f_s is the sampling frequency and $\lfloor \cdot \rfloor$ is the floor operation. The decimation reduces the number of samples to process and thus reduces the computational cost. The output of this stage is an $N_{TF} \times N_R$ matrix with $N_R = \lfloor \frac{N}{R} \rfloor$. The decimated subband signals $x_{R,i}(n)$ are then logarithmically compressed and each frame is normalized to unit variance leading to the output signals $\mathbf{y}_{R,i} = [y_{R,i}(1)\, y_{R,i}(2)\, \ldots\, y_{R,i}(N_R)]$, which are arranged into a matrix $\mathbf{Y}_{TF} = \left[\mathbf{y}_{R,1}^T\, \mathbf{y}_{R,2}^T\, \ldots\, \mathbf{y}_{R,N_{TF}}^T\right]^T$. The second filter bank block in Figure 1 contains for each subband signal $\mathbf{y}_{R,i}$, with $i \in \{1, \ldots, N_{TF}\}$, N_M parallel recursive filters and thus contains $(N_{TF} \times N_M)$ filters in total. The output signals of these filters are denoted by $y_{R,i,j}(n)$ with $j \in \{1, \ldots, N_M\}$. The ordering of the cutoff frequencies is similar to the first filter bank as depicted in Figure 2. In the next two steps the $(N_{TF} \times N_M)$ subband signals are rectified, averaged over the frame length and stored in the vector $\mathbf{Y}_{MF} = \left[y_{MF,1}^T\, y_{MF,2}^T\, \ldots\, y_{MF,N_{TF} \times N_M}^T\right]^T$ with

$$y_{MF,((i-1) \cdot N_M + j)} = \frac{1}{N_R} \sum_{n=1}^{N_R} |y_{R,i,j}(n)| \quad . \qquad (1)$$

In this work we set the frame duration $(N_R \cdot R)/f_s$ to 2s. The feature vector $\mathbf{Y}_{MF}$ is then fed into a classification model or regression model. Because of the fact that this special structure of successive filter banks extract spectral amplitude modulations of the input signal, we call these features AMS (amplitude modulation spectrum) features. In this AMS feature extraction the number of tunable parameters that have to optimized, is $N_{TF} - 1 + (N_{TF} \times N_M)$, namely the bandwidths of the filters. Any filter design rule could be used for the filters. In this paper, we use the Butterworth filter design for the filters in the first filter bank and the Chebyshev II filter design rule for the filters in the second filter bank.
The bandwidths of these filters are found by a model-based optimization (MBO) algorithm which is described in the next section.

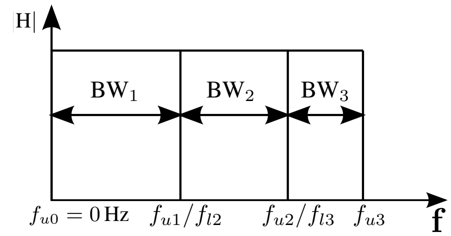

Fig. 2: Filter band structure for three filters. f_{li} is the lower edge frequency of the i-th filter, f_{ui} is the upper edge frequency of the i-th filter, and BW_i is the bandwidth of the i-th filter

B. Model-Based Optimization

Next we describe the training approach of the filter banks used for the AMS feature extraction in more detail. Using the AMS features and a classifier we minimize the mismatch between the estimated class labels and the ground truth class labels. The tunable parameters are the bandwidths of the filters in the filter banks. In the following we will give a brief outline of the main idea of the used MBO method. For a more detailed description of MBO we recommend [11], [12].
MBO, like the Covariance Matrix Adaptation Evolution Strategy (CMA-ES [13]) which we used in our previous work [5], is an iterative approach used to optimize a black box objective function. It is used when the evaluation of an objective function, in our case the classification error depending on different filter bank parameters, is expensive in terms of available resources (computational time). MBO constructs an approximation model, a so called surrogate model, of this expensive objective function to find the optimal parameters for a given problem. The evaluation of the surrogate model is less complex than the original objective function. In initial experiments, we found that MBO is roughly 1.6 times faster than the CMA-ES implementation and also resulted in better solutions for previously investigated acoustic scene classification tasks, thus motivating its use in this work.
We can divide MBO in three steps, which will be explained in the following three subsections.

1) Design a Sampling Plan: We assume a high dimensional multi-modal parameter space. The goal of the optimization is to find the point, which minimizes the cost function. The initial step of the MBO is to construct a sampling plan. This means that we determine n points which will then be evaluated by the objective function. These n points should cover the whole parameter space and for this we use the Latin hypercube design [11].

2) Constructing a Surrogate Model: The surrogate model $\hat{g}(\mathbf{x})$ should be constructed such that it is a reasonable approximation of the objective function $y = f(\mathbf{x})$ where $\mathbf{x}$ is a k-dimensional parameter vector. Among available models [11] we use the ordinary kriging model in this paper:

$$\hat{g}(\mathbf{x}) = \mu + Z(\mathbf{x}) \tag{2}$$

where μ can be interpreted as a constant global mean and $Z(\mathbf{x})$ is a Gaussian process. The mean of this Gaussian process is 0 and its covariance is

$$\mathrm{Cov}(Z(\mathbf{x}), Z(\mathbf{x})) = \sigma^2 \rho(\mathbf{x} - \mathbf{x}', \Psi) \tag{3}$$

with ρ the *Matern 3/2* kernel function and Ψ a scaling parameter. The constant σ^2 can be interpreted as the global variance. Thus, the unknown parameters of this model are μ, σ^2 and Ψ. We have to estimate these parameters by using the n previously evaluated points $\mathbf{y} = (y_1, ..., y_n)^T$.
The likelihood function is given by

$$L(\mathbf{y}; \mu, \sigma^2, \Psi) = \frac{1}{\sqrt{(2\pi)^n \sigma^{2n} det(R)}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{1}\mu)^T R^{-1}(\mathbf{y} - \mathbf{1}\mu)\right) \tag{4}$$

with $R = (\rho(\mathbf{x}_i - \mathbf{x}_j, \Psi))_{i,j=1,...,n}$ and $det(R)$ its determinant. From this we can determine the maximum likelihood estimation of the unknown parameters μ, σ^2 and Ψ.

3) Exploring and Exploiting a Surrogate Model: One can derive the surrogate prediction $\hat{f}_n(\mathbf{x})$ and the corresponding prediction uncertainty $\hat{s}_n(\mathbf{x})$ (see [11] [12]) based on the first n evaluations of $f(\mathbf{x})$. The estimated surrogate function follows a normal distribution $\hat{g}(\mathbf{x}) \sim \mathcal{N}\left(y; \hat{f}_n(\mathbf{x}), \hat{s}_n^2(\mathbf{x})\right)$.
Let the actual best value be $y_{\text{min}} = \min_{i=1,...,n} y_i = \min_{i=1,...,n} f(\mathbf{x}_i)$. The improvement for a point $\mathbf{x}$ and the estimated surrogate $\hat{g}(\mathbf{x})$ is $I_n(\mathbf{x}) = \max(y_{\text{min}} - \hat{g}(\mathbf{x}), 0)$. We find the next point to evaluate by maximizing the expected improvement $\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}} \mathrm{E}(I_n(\mathbf{x}))$. This criterion gives us a balance between exploration (improving global accuracy of the surrogate model) and exploitation (improving local accuracy in the region of the optimum of the surrogate model). In this way the optimizer does not get stuck in local minima and converges to a better solution.
After each iteration of MBO the surrogate model will be updated. Different convergence criteria could be chosen to determine when to stop evaluating new points for updating the surrogate model, after e.g., a preset number of iterations (here 200), or after the expected improvement drops below a predefined threshold.

C. Classification and Regression

As described in the previous section the MBO is minimizing the cost of an objective function. In order to optimize the filter banks, we chose, due to its low computational complexity and the assumption of Gaussian class densities, the accuracy of a linear discriminant analysis (LDA) classifier as the cost function. The LDA does not need any parameter to be tuned, which simplifies the optimization procedure. In general, any kind of classification or regression algorithm could be used as well and also other cost functions (e.g. cross entropy) could be chosen. After optimizing the filter bandwidths we replace the LDA with a partial least squares (PLS) regression and trained a PLS model with the optimized filter banks on the same training set. Due to the restriction on the filter passbands (see Figure 2) no gaps in the frequency spectrum are possible and thus some filter passbands might not be useful and can be discarded. In order to find the best feature combination we evaluated all possible combinations and chose the one with the best estimation results on the training set. The so optimized filters, the determined feature set and the trained PLS model were then used for the evaluation of unseen data samples.

III. Results

A. Data

In this paper we evaluate our approach on the ACE challenge corpus [14]. The task of this challenge was to estimate the reverberation time T_{60} and the direct-to-reverberant ratio from recorded audio files.
The ACE corpus is composed of two sets: a development set DEV for training and testing an estimator and an evaluation set EVAL for evaluating the final estimator. Both sets can be freely downloaded from the ACE challenges web site [14]. These sets are created with recorded acoustic impulse responses (AIR) with different microphone configurations which are then rendered with different speakers, three noise types (ambient, fan and babble) and different SNR levels. For the DEV set AIRs from two different rooms and two different room positions are used, leading to four different DRR values (Table I) and two T_{60} times. The SNR levels are set to 0 dB, 10 dB, and 20 dB. The total number of files is 288.
For the EVAL set AIRs from five different rooms and two different room positions are used, leading to 10 different DRR values (Table II). The SNR levels are set to -1 dB, 12 dB and 18 dB. The total number of files is 4500.
The corpus provides audio signals for a number of different microphone arrays and setups, from which we used the single channel input with a sampling frequency of 16 kHz at a resolution of 16bit.
The quality or comparison measures used in the ACE challenge are: the bias (mean estimation error), mean squared error MSE, Pearson correlation coefficient between ground truths and estimations ρ and the real-time factor RTF, which is defined as the computation time divided by the duration of processed speech files. Additionally, we provide the root mean squared error (RMSE). We run our experiments on a PC with an Intel(R) Core(TM) i5-3470 CPU @ 3.20 GHz with 12GB RAM and Matlab 2017b.

Room	Position	DRR [dB]	T_{60} [s]
Building Lobby	1	8.10	0.7389
Building Lobby	2	5.13	0.7655
Office 1	1	10.45	0.3513
Office 1	2	4.34	0.3103

TABLE I: ACE challenge corpus DEV set and ground-truth parameters [14].

Room	Position	DRR [dB]	T_{60} [s]
Lecture Room 1	1	12.20	0.6232
Lecture Room 1	2	5.09	0.6657
Lecture Room 2	1	8.94	1.3321
Lecture Room 2	2	4.96	1.2926
Meeting Room 1	1	1.71	0.4501
Meeting Room 1	2	0.94	0.4447
Meeting Room 2	1	5.00	0.3804
Meeting Room 2	2	8.83	0.3831
Office 2	1	2.44	0.4065
Office 2	2	-2.28	0.3985

TABLE II: ACE challenge corpus EVAL set and ground-truth parameters [14].

#Filters	Bias [s]	MSE	RMSE [s]	ρ
3x2	0.00	0.0151	0.123	0.814
3x3	0.00	0.0287	0.170	0.598
4x2	0.00	0.0337	0.184	0.497
4x4	0.00	0.0133	0.116	0.838
4x5	**0.00**	**0.0131**	**0.115**	**0.840**
5x5	0.00	0.0267	0.166	0.619

TABLE III: T_{60} estimation results on the DEV set of the ACE challenge corpus with different filter configurations.

#Filters	Bias [s]	MSE	RMSE [s]	ρ
3x2	-0.418	0.301	0.548	0.141
3x3	0.0969	0.145	0.381	0.0425
4x2	0.135	0.138	0.371	0.2047
4x4	0.142	0.137	0.370	0.256
4x5	0.104	0.119	0.345	**0.390**
5x5	**0.0471**	**0.116**	0.**341**	0.271

TABLE IV: T_{60} estimation results on the EVAL set of the ACE challenge corpus with filter configurations optimized on the DEV set.

B. Filter Bandwidths Obtained from MBO

We use MBO to optimize different filter bank configurations as depicted in Table III on the DEV set. For the T_{60} estimation task the best filter configuration was found for $N_{TF} = 4$ time-domain filters and $N_M = 5$ modulation domain filters which all of are sixth-order IIR filters. Thus, this 4×5 filter bank configuration extracts 20 features for each 2s segment. In order to find the most contributing and also the best features leading to the best performance on the DEV set, we evaluate all 2^{20} possible feature combination. This brute-force search determined five features leading to the best estimation result. The MBO of $N_{TF} = 4$ and $N_M = 5$ filter bandwidths took $\sim$ 8h for 200 iterations.

Similar to T_{60} estimation, we optimized different filter configurations for the DRR estimation task. The best filter configuration on the DEV set was obtained with $N_{TF} = 3$ time-domain filters and $N_M = 2$ modulation domain filters, which all off are sixth-order IIR filters. With this 3×2 filter configuration we extracted 6 features for each 2s segment. We evaluated all 2^6 possible feature combination and found that five features contributed to the best DRR estimation result. The MBO of $N_{TF} = 3$ and $N_M = 2$ filter bandwidths took $\sim$ 5h for 200 iterations.

C. Partial-Least Square Regression

A PLS regression model was trained on the DEV data with the best filter parameters obtained from the MBO and applied on the EVAL data. The results for T_{60} and DRR estimation on the EVAL set are shown in Table IV and Table V. Among the submitted approaches in the T_{60} estimation task the proposed approach ranks in the upper mid-range with an RMSE of 0.345s, a bias of 0.104s and a correlation coefficient $\rho = 0.390$ [14]. Due to the low computational complexity of our algorithm, we achieve a RTF value of 0.00469. The T_{60} estimation error is depicted in Figure 3. It can be seen that the two both highest T_{60} conditions have the largest absolute errors. A reason for this could be that such values were not present in the DEV set (see Table I).

Figure 4 depicts the DRR estimation on the EVAL set. Obviously, for low DRR values the estimation error is positive and for higher DRR values negative. The best estimation accuracy is obtained for DRR values around 5 dB. Compared to results submitted to the ACE challenge for DRR estimation we rank third best in terms of MSE. The corresponding RMSE is 3.74 dB. In comparison to other submitted algorithms it also has a low bias of -0.199dB, the sixth best correlation coefficient with $\rho = 0.462$, and the best RTF with 0.0027 as shown in Table V. Our algorithm scores better than all other submitted algorithms using the single channel microphone configuration for this estimation task in terms of MSE, bias, and RTF values.

IV. Discussion and Conclusion

In this paper we show that the presented AMS-based feature extraction algorithm, which was used for acoustic scene classification and source localization before [5], [6], can also be used for T_{60} and DRR estimation tasks. For both tasks we find that the estimation performance depends on the filter bank configuration. For instance the DRR estimation works better with a lower number of filters and the T_{60} estimation with a slightly larger configuration. The filter banks needed for the AMS feature extraction are optimized by an MBO algorithm on the DEV set.

We evaluated our proposed approach for T_{60} and DRR estimation on the publicly available ACE challenge corpus in order to compare it with other proposed algorithms. With an RMSE of 3.74 dB for DRR estimation we outperform the

#Filters	Bias [dB]	MSE	RMSE [dB]	ρ	RTF
3x2	-0.199	13.99	3.74	0.462	0.00274

TABLE V: DRR estimation results on the EVAL set of the ACE challenge corpus.

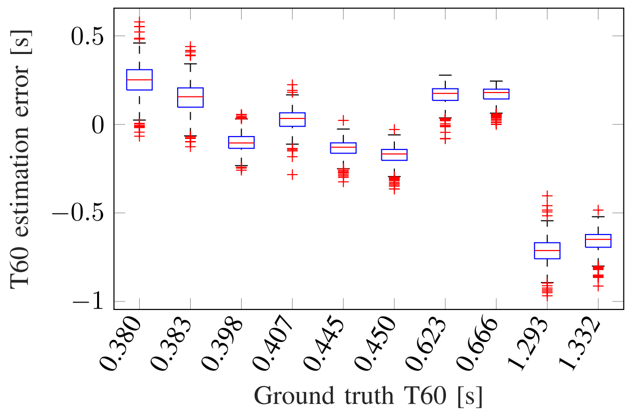

Fig. 3: T_{60} estimation error on the EVAL set of the ACE challenge corpus.

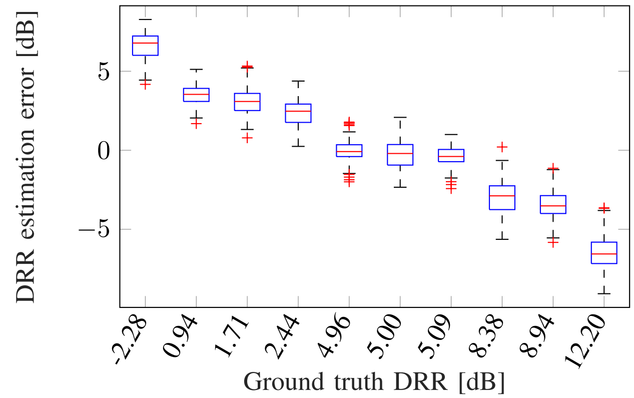

Fig. 4: DRR estimation error on the EVAL set of the ACE challenge corpus.

submitted algorithms based on neural networks (3.83 dB [3] and 3.99 dB [15]) and all single-channel algorithms and obtain the third best RMSE value with only 5 features. However, the correlation coefficient of 0.462 is slightly lower than the neural-network-based algorithms (0.558 [3] and 0.556 [15]), and thus is the fourth best correlation coefficient among the algorithms using the single channel microphone configuration. For T_{60} estimation we rank in the upper mid-range of performance with an RMSE of 0.345 s and a correlation coefficient of 0.39 with just 5 features. The RTF outperforms all other submitted algorithms in both estimation tasks significantly.
Compared to other known AMS-based feature extraction algorithms [4], [7]–[10] our approach is computationally significantly less complex. The low computational complexity makes our AMS-based approach suitable for low-resource and also real-time systems. In this paper we also show that the AMS-based feature extraction framework can be easily adapted for the estimation of room acoustic parameters with its regular but flexible structure. Similarly to deep-neural-network-based structures determining the best filter bank configuration can be a challenging task and it depends on the given application. However, because of the low number of parameters a global optimization approach such as MBO can be employed.

References

[1] T. de M. Prego, A. A. de Lima, R. Zambrano-Lpez, and S. L. Netto, "Blind estimators for reverberation time and direct-to-reverberant energy ratio using subband speech decomposition," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.

[2] H. W. Löllmann, A. Brendel, P. Vary, and W. Kellermann, "Single-channel maximum-likelihood T60 estimation exploiting subband information," *CoRR*, vol. abs/1511.04063, 2015. [Online]. Available: http://arxiv.org/abs/1511.04063

[3] P. P. Parada, D. Sharma, T. van Waterschoot, and P. A. Naylor, "Evaluating the non-intrusive room acoustics algorithm with the ACE challenge," *CoRR*, vol. abs/1510.04616, 2015. [Online]. Available: http://arxiv.org/abs/1510.04616

[4] F. Xiong, S. Goetze, B. Kollmeier, and B. T. Meyer, "Exploring auditory-inspired acoustic features for room acoustic parameter estimation from monaural speech," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1809–1820, Oct 2018.

[5] S. Ağcaer, A. Schlesinger, F. Hoffmann, and R. Martin, "Optimization of amplitude modulation features for low-resource acoustic scene classification," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 2556–2560.

[6] S. Ağcaer and R. Martin, "Binaural source localization based on modulation-domain features and decision pooling," *CoRR*, vol. abs/1812.02399, 2018. [Online]. Available: http://arxiv.org/abs/1812.02399

[7] J. Anemüller, D. Schmidt, and J. Bach, "Detection of speech embedded in real acoustic background based on amplitude modulation spectrogram features," in *Interspeech 2008 9th Annual Conference of the International Speech Communication Association*. Interspeech, 2008, pp. 2582–2585.

[8] J. Bach, J. Anemüller, and B. Kollmeier, "Robust speech detection in real acoustic backgrounds with perceptually motivated features," *Speech Communication*, vol. 53, no. 5, pp. 690 – 706, 2011, perceptual and Statistical Audition. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639310001305

[9] G. Langner, M. Sams, P. Heil, and H. Schulze, "Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography." *J Comp Physiol A*, vol. 181, no. 6, pp. 665 – 676, 1997.

[10] N. Moritz, J. Anemüller, and B. Kollmeier, "Amplitude modulation spectrogram based features for robust speech recognition in noisy and reverberant environments," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 5492–5495.

[11] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, September 2008. [Online]. Available: http://amazon.com/o/ASIN/0470060689/

[12] C. Weihs, S. Herbrandt, N. Bauer, K. Friedrichs, and D. Horn, "Efficient global optimization: Motivation, variations, and applications," in *Archieves of Date Science*, 2016.

[13] N. Hansen, "The CMA evolution strategy: A comparing review," in *Towards a new evolutionary computation. Advances in estimation of distribution algorithms*. Springer, 2006, pp. 75–102.

[14] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: The ace challenge," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1681–1693, Oct 2016.

[15] F. Xiong, S. Goetze, and B. T. Meyer, "Joint estimation of reverberation time and direct-to-reverberation ratio from speech using auditory-inspired features," *CoRR*, vol. abs/1510.04620, 2015. [Online]. Available: http://arxiv.org/abs/1510.04620