

# ROOM EQUALIZATION BASED ON MEASUREMENTS WITH MOVING MICROPHONES

*Radoslaw Mazur, Fabrice Katzberg, Huy Phan, and Alfred Mertins*

Institute for Signal Processing  
University of Lübeck  
Ratzeburger Allee 160, 23562 Lübeck, Germany

## ABSTRACT

For enhancing the perceived quality of a sound played in a closed room, the methods of room impulse response equalization are used. For a better control of late echoes, new approaches, such as temporal masking, exploit the properties of the human auditory system. In order to allow the listener to move freely even in a small area, a whole volume has to be equalized. Traditional methods require a measurement of a huge amount of impulse responses in this volume. In this work we propose a method which allows for a greatly reduced measurement effort. Here, we employ a dynamic method for measurement of impulse responses using just one moving microphone. The reconstructed impulse responses are used to perform equalization and a simple interpolating technique allows for a equalization at the position of the listener, who can freely move inside the measured volume.

*Index Terms*— Room Impulse Response, Reshaping, Equalization, Moving Microphones

## 1. INTRODUCTION

In typical hands free communication systems, sounds are played using loudspeakers. In a closed room, the added reverberation, as described by the room impulse response (RIR), usually results in a degraded perceived quality. In order to combat this problem, the played sound can be prefiltered in a way, that the global impulse response (GIR, the convolution of the prefilter and the RIR) has no audible influence on the signal [1].

Early approaches try to minimize a distance of the GIR to a shifted dirac pulse or, more general, a bandpass in a quadratic sense [2, 3]. Although these methods are successful in minimizing the unwanted parts of the GIR, the audible results are not satisfactory due to perceived late echoes. Therefore, in recent works, the properties of the human auditory system had been taken into account. The aim there is to create a prefilter which is not an inverse of the RIR, but will just render the echos inaudible. This relaxed requirement with the use of the average temporal masking curve [4] has been

very successful [5]. For a better control of the late echoes the method has been modified to use a  $p$ -norm based criterion instead of the quadratic term. There are more extensions for this approach; for example, an explicit control of the frequency domain [6].

Unfortunately, when aiming for a hassle-free hands free communication, these approaches are not sufficient. Small movements of the listener result in a changed RIR and the performance of the setup degrades [7]. For a more spatially robust method, different approaches have been proposed. The algorithms can be generally sorted into two classes. One approach is to use the multi-position method. Here, with the knowledge of multiple RIRs inside the listening area, the pre-filters are designed to equalize multiple points in space [8, 9]. When the sampling points suffice the time-space sampling theorem [9] (i. e. are dense enough), the whole volume is equalized. In order to get satisfactory results for bigger volumes, multiple loudspeakers have to be used. There are serious drawbacks with this MIMO approach, as it requires us to measure all the RIRs from all loudspeakers to all positions in the listening area. The practical implementation of these measurements is very time consuming and may even require a calibrated array of microphones and is not feasible for any real-world application. The needed calculations are also very demanding. A CUDA implementation on graphics cards has been proposed in [10]. The other approach is to model additional errors in the optimization and adding regularizers [8, 11]. In [12], a method has been proposed to generate hypothetical RIRs instead of measured ones. These approaches are a trade-off, as the extended equalized volume comes at the cost of a worse performance at the center point.

In this work we propose a new approach to achieve reshaping in a whole volume. Instead of estimating the RIRs independently, we employ a dynamic measurement for the whole listening area using moving microphones. The method proposed in [13, 14] uses perfect sequences and allows almost arbitrary trajectories in the listening area. With spatial interpolation [15] the recovery of a dense grid of RIRs is possible. Even in the simplest case with one loudspeaker and one microphone this method allows for a very fast estimation of the sought RIRs, which can be used to calculate the equalizers.

The paper is organized as follows. In the next section the

This work has been supported by the German Research Foundation under Grants No. ME 1170/8-1 and ME 1170/10-1.

basics of the dynamic sampling procedure will be discussed. Then, in Section 3, an equalization method will be revised. In Section 4, results of the simulations will be shown. Finally, some conclusions will be given in Section 5.

## 2. DYNAMIC MEASUREMENT OF RIRS

For the description of the spatio-temporal sound field in the listening area, the plenacoustic function (PAF) has been introduced in [15]. Without loss of generality, with a fixed source at a position the PAF becomes the spatio-temporal RIR  $p(\mathbf{r}, n) = c(\mathbf{r}, n)$  with  $\mathbf{r}$  being the position in the 3D space [13]. The traditional method of obtaining  $c(\mathbf{r}, n)$  is a dense sampling on a equidistant grid

$$\mathcal{G} = \{\mathbf{r}_g | \mathbf{r}_g = \mathbf{r}_0 + [g\Delta_x, g\Delta_y, g\Delta_z]\} \quad (1)$$

with some position  $\mathbf{r}_0$ , the sampling intervals  $\Delta_\xi$  in the three dimensions, and  $g \in \mathbb{Z}$ . According to the Nyquist-Shannon sampling theorem, for a sampling frequency  $f_c$ , the temporal condition  $f_s > 2f_c$ , and the spatial condition

$$\Delta_\xi < \frac{c_0}{2f_c}, \quad \xi \in x, y, z \quad (2)$$

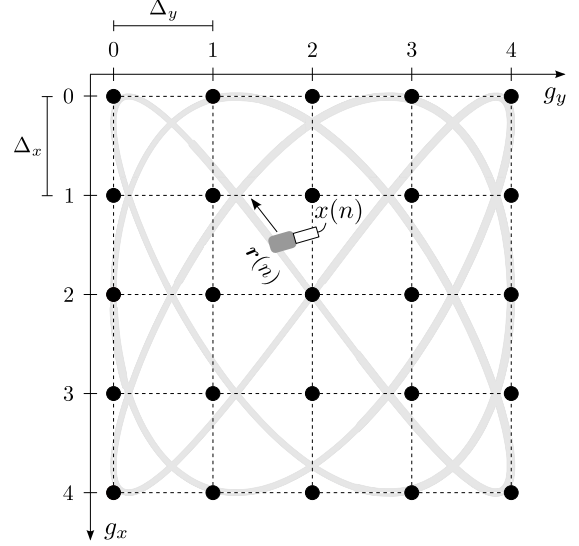
with  $c_0$  being the speed of sound has to be fulfilled. For practical cases this requires a huge effort. When using a single microphone, these points have to be sampled successively which requires a lot of time. When using an array of microphones, this time can be reduced, but usually at the cost of microphone calibration and possibly higher effort for a precise repositioning of the array.

In [13] an alternative method using moving microphones has been proposed. Here, one source at a fixed position emits a signal  $s(n)$ , which is recorded by a microphone moving in the listening area. The trajectory can be arbitrary, as long as it covers the whole area, and has to be known or recorded during the procedure. Note that an extension for the use of an array of moving microphones is simple and allows for a trade-off between calibration costs and measurement time.

In Fig. 1 a sketch of such a scenario in a 2D case is shown. Here, the reference point  $\mathbf{r}_0$  is in the upper left corner. The grid size is  $5 \times 5$ . The microphone is moved on a Lissajous trajectory which covers the area defined by the grid. The use of the Lissajous trajectory has some advantages, as it has a mostly constant velocity and covers the area in almost uniform manner, being only little more dense on the edges. In practice, however, a handheld microphone with position tracking can be used instead.

The recorded signal  $x(n)$  is a convolution of the source signal with the RIR  $c(\mathbf{g}, n)$  at the position  $\mathbf{g}$  of the microphone at time point  $n$ . Alternatively, with an interpolation function  $\varphi(\mathbf{r}(n), \mathbf{r}_g)$  between a RIR at a grid point and a displacement  $\mathbf{r}(n) - \mathbf{r}_g$ , the recorded signal can be described as

$$x(n) = \sum_{k=0}^{L-1} \sum_{g \in \mathcal{G}} \varphi(\mathbf{r}(n), \mathbf{r}_g) c(\mathbf{r}, k) s(n-l) + \eta(n) \quad (3)$$



**Fig. 1.** The dynamic sampling procedure: Arrangement of a 2D sampling grid with the RIRs being reconstructed at the solid dots. In gray, a Lissajous trajectory for the dynamic microphone is shown.

with  $L$  being the length of the RIR and  $\eta(n)$  the measurement noise. In order to solve (3) for the sought RIRs, we concatenate the  $N = XYZ$  grid RIRs of length  $L$  in a vector  $\mathbf{c} \in \mathbb{R}^{NL}$  and define  $\mathbf{x} = [x(0), \dots, x(M-1)]^T$  and the noise vector  $\boldsymbol{\eta}$ . This leads to a set of linear equations

$$\mathbf{x} = \mathbf{A}\mathbf{c} + \boldsymbol{\eta}. \quad (4)$$

The matrix  $\mathbf{A}$  has block structure

$$\mathbf{A} = [\boldsymbol{\Phi}_1 \mathbf{S}, \boldsymbol{\Phi}_2 \mathbf{S}, \dots, \boldsymbol{\Phi}_N \mathbf{S}] \quad (5)$$

with  $\boldsymbol{\Phi}_u \in \mathbb{R}^{M \times M}$  being a diagonal matrix concatenating the interpolation coefficients of all  $M$  measurements for the  $u$ -th RIR on the grid  $\mathcal{G}$ .  $\mathbf{S}$  is the convolution matrix of the source signal.

Due to the size of  $\mathbf{A}$ , a direct solution of (4) is not feasible. Fortunately, when using a pseudo random excitation signal with perfect autocorrelation [16, 17], (4) can be reformulated to  $\mathbf{A}$  having a block structure [13]. This allows a decoupling of the time dimension from  $\mathbf{A}$ . This transformation reduces the problem into  $L$  smaller problems, with  $1/L$  number of unknowns. Since these problems can be solved independently, this huge reduction of the computational cost makes the whole procedure applicable to bigger grids.

## 3. RIR RESHAPING

For the reshaping method from [8] the RIRs  $c^{(i)}(n)$  of length  $L_c$  from a loudspeaker to the  $i$ -th position in space have to be known. With the prefilter  $h(n)$  of length  $L_h$ , the overall impulse responses are given by  $g^{(i)}(n) = h(n) * c^{(i)}(n)$ . The

reshaping is carried out accordingly to the desired and unwanted parts of the RIR, which are defined using the windows  $w_d(n)$  and  $w_u(n)$ . The desired part is given by  $g_d^{(i)}(n) = g^{(i)}(n)w_d(n)$  and analogously for the unwanted part.

The prefilter is obtained by solving the optimization problem given by

$$\text{MIN}_{\mathbf{h}} : f(\mathbf{h}) = \log \left( \frac{f_u(\mathbf{h})}{f_d(\mathbf{h})} \right) \quad (6)$$

with

$$f_d(\mathbf{h}) = \|\mathbf{g}_d\|_{p_d} = \left( \sum_{i=1}^{N_m} \sum_{k=0}^{L_g-1} |g_d^{(i)}(k)|^{p_d} \right)^{\frac{1}{p_d}} \quad (7)$$

and  $f_u(\mathbf{h}) = \|\mathbf{g}_u\|_{p_u}$ , accordingly.  $N_m$  is the number of target points in the listening area. The vectors  $\mathbf{g}_d$  and  $\mathbf{g}_u$  consists of stacked wanted and unwanted parts of the  $N_m$  global RIRs. For the solution a gradient based optimization is used [8].

In contrast to the least-squares methods [2, 3], with large values for  $p_d$  and  $p_u$  (typically between 10 and 20), a very smooth shaping, with no outliers, can be achieved.

#### 4. EXPERIMENTS AND RESULTS

The experiments were performed on simulated data, so the ground truth was available. Here, we used the same setup as in [13], which means a listening area of  $0.12\text{m} \times 0.12\text{m}$  in a typical sized office room. The RIRs were simulated using the shoe box model as in [18]. The lengths of the RIRs were 1000 taps at a sampling frequency of 8 kHz. the excitation signal was an MLS sequence of length 1023. The measurement trajectory was a Lissajous figure with a ratio of 33/32 covering the whole listening area. The reconstruction was performed using a Lagrange interpolator. A direct solution of (4) for this setup would require to solve at once for  $1.69 \cdot 10^5$  unknowns. With the decoupling method from [13] this problem was reduced to one thousand of independent  $169 \times 169$  systems. To simulate real measurements, the additive noise in (3) was set to be white and have an SNR of 40 and 60dB respectively. These values represent a typical and a quiet office room.

With a grid size of 1cm, 169 RIRs have been recovered. In the following, two experiments using these RIRs will be discussed. In the first one, the general feasibility of the measurement method will be tested. In the second experiment the spatial robustness will be examined.

The quality of reconstruction will be evaluated using the nPRQ measure form [8], which calculates the overshoot above the temporal masking curve being above  $-60\text{dB}$  of the main peak  $g_{os}(n) = \max \left( \frac{1}{w_u(n)}, -60\text{dB} \right)$  as

$$\text{nPRQ} = \begin{cases} \frac{1}{\|\mathbf{g}_E\|_0} \sum_{n=N_0}^{L_g-1} g_E(n), & \|\mathbf{g}_E\|_0 > 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

**Table 1.** Comparison of the results for different measurement noise.

| RIR                           | nPRQ  |
|-------------------------------|-------|
| Reference                     | 11.25 |
| Reference reshaped            | 8.59  |
| Reconstructed (60dB) reshaped | 8.65  |
| Reconstructed (40dB) reshaped | 9.56  |

with

$$g_E(n) = \begin{cases} 20 \log_{10}(|g(n)|w_u(n)), & |g(n)| > g_{os}(n) \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

Lower values indicate less reverberation, and a value of zero means no audible echoes.

#### 4.1. Experiment 1

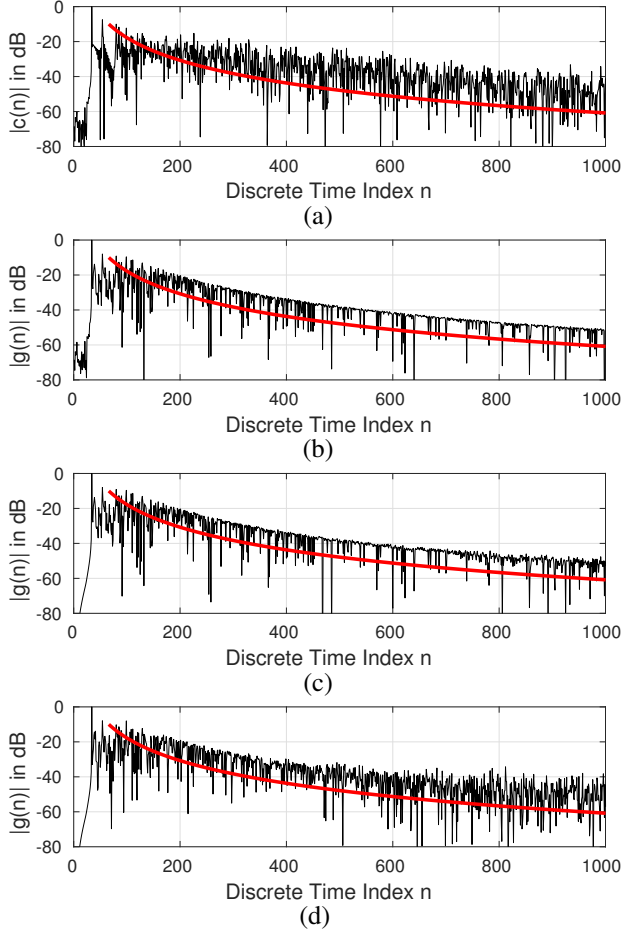
The first experiment was performed on single points in the listening area. Here, the reconstructed RIRs for both noise levels are compared to the reference. In Fig. 2, a typical situation is shown. For reference, in Fig. 2 (a), the measured RIR is shown in black on a logarithmic scale with the temporal masking curve in red. In this situation we have quite high reverberation with a lot of audible echoes. Note that in [6, 8, 12], usually lower levels of reverberation have been used. The reconstructed RIRs for both noise levels (60dB and 40dB) are reconstructed with a relative error of around 41dB and 21dB, respectively. Even on the logarithmic scale they look virtually the same and are not shown explicitly.

In Fig. 2 (b) the overall RIR  $g(n)$  is shown for the case when the equalizer is designed using the reference RIR. The reshaping was able to lower the nPRQ measure by about 2.7 as shown in Table 1. The overshoot above the temporal masking curve is roughly the same for all time points. The audible result is less reverberation.

In Fig. 2 (c) the equalizer has been designed on the measured RIR with the noise level of 60dB. The graph shows the result when applied on the reference RIR. While the overshoot is not as smooth as in the previous case, the reshaping is still successful. The reduction of the nPRQ is 2.6 and almost as good as for the reference.

In Fig. 2 (d) the result for the noise level of 40dB is shown. Here, the reshaping is still able to lower the reverberation, but the overshoot is not smooth and is higher than in the previous case. The improvement in terms of nPRQ is only 1.7.

This experiment confirms, that the new dynamic measurement of RIRs is able to provide RIRs which can be used for reshaping. In order to get satisfactory results a good control of the measurement noise is needed.

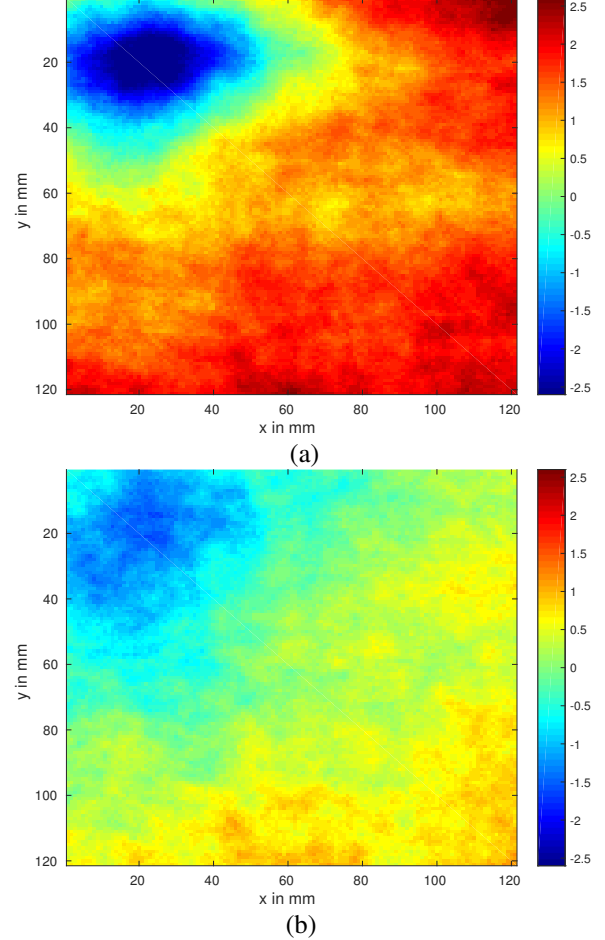


**Fig. 2.** Reshaping results for a single point. (a) The original RIR at a central position of the listening area in logarithmic representation. The reconstructed RIRs look virtually the same on this scale and therefore are not shown. (b) The overall RIR when reshaping using the true RIR. (c) and (d) The overall RIR when using the reconstructed RIR at noise levels of 60dB and 40dB.

## 4.2. Experiment 2

In the second experiment we examine the spatial robustness. At first, we again calculate an equalizer for a single point. By applying it to the whole volume, we can examine the improvement and deterioration inside the listening area. In Fig. 3 (a) the results are shown. Here, the image shows a color coded difference to the non-equalized volume. Blue indicates improvement; green shows no change, and red color means deterioration. This single channel result is comparable with previous works, and again indicates the eligibility of the new measurement method.

In Fig. 3 (b) an array of  $5 \times 5$  RIRs, which corresponds to an area of  $4\text{cm} \times 4\text{cm}$ , has been equalized. Here, the equalization is not as successful in terms of nPRQ, but the equalized area has been significantly increased. Additionally, the dete-



**Fig. 3.** Equalization of a volume. In (a) a single RIR has been reshaped. In (b) a grid of  $5 \times 5$  points has been used. The color codes the improvement/deterioration of the perceived echoes in terms of nPRQ.

rioration in the other parts has been reduced.

As the extended area comes at the cost of worse performance at the center, it is not feasible to equalize the whole listening area, when there is only one loudspeaker available. Still, when tracking is available, as it is already needed for the measurement with the dynamic microphone, the area of interest can be freely moved in the listening area by using precalculated equalizers for the different positions.

## 5. SUMMARY

In this work we combined a new dynamic method for measurement of impulse responses using one moving microphone with the computation of equalizers based on the properties of the human auditory system. With this simple setup of one loudspeaker and one microphone, arbitrary parts of the listening area could be equalized. With precalculated equalizers and an appropriate tracking system, this allows for head movements in a hands free communication system.

## 6. REFERENCES

- [1] John N. Mourjopoulos, "Digital equalization of room acoustics," *Journal of the Audio Engineering Society*, vol. 42, no. 11, pp. 884–900, Nov. 1994.
- [2] Stephen J. Elliott and Philip A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," *Journal of the Audio Engineering Society*, vol. 37, no. 11, pp. 899–907, Nov. 1989.
- [3] Markus Kallinger and Alfred Mertins, "Room impulse response shortening by channel shortening concepts," in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct. 2005, pp. 898–902.
- [4] Louis D. Fielder, "Practical limits for room equalization," in *Proc. 111th Convention of the Audio Engineering Society*, Nov. 2001, pp. 1–19.
- [5] Alfred Mertins, Tiemin Mei, and Markus Kallinger, "Room impulse response shortening/reshaping with infinity- and p-norm optimization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 249–259, Feb. 2010.
- [6] Jan Ole Jungmann, Stefan Goetze, and Alfred Mertins, "Room impulse response reshaping by p-norm optimization based on estimates of room impulse responses," in *Proceedings of the German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, Mar. 2011, pp. 611–612.
- [7] Biljana D. Radlović, R. C. Williamson, and Rodney A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, May 2000.
- [8] Jan Ole Jungmann, Radoslaw Mazur, Markus Kallinger, Tiemin Mei, and Alfred Mertins, "Combined acoustic mimo channel crosstalk cancellation and room impulse response reshaping," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1829–1842, Aug. 2012.
- [9] Tiemin Mei and Alfred Mertins, "On the robustness of room impulse response reshaping," in *Proceedings of the IEEE International Workshop on Acoustic Echo and Noise Control*, Tel Aviv, Israel, Aug. 2010.
- [10] Radoslaw Mazur, Jan Ole Jungmann, and Alfred Mertins, "On cuda implementation of a multichannel room impulse response reshaping algorithm based on p-norm optimization," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, Oct. 2011, pp. 305–308.
- [11] Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins, "A stochastic approach for robust listening room compensation," in *Proceedings of the German Annual Conference on Acoustics (DAGA)*, Merano, Italy, Mar. 2013, pp. 185–186.
- [12] Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins, "Perturbation of room impulse responses and its application in robust listening room compensation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, BC, Canada, May 2013, pp. 433–437.
- [13] Fabrice Katzberg, Marco Maass, Radoslaw Mazur, Philipp Koch, and Alfred Mertins, "Sound-field measurement using moving microphones," *Submitted*, 2016.
- [14] Fabrice Katzberg, Radoslaw Mazur, Marco Maass, Philipp Koch, and Alfred Mertins, "Measurement of sound fields using moving microphones," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2017, *Accepted*.
- [15] Thibaut Ajdler, Luciano Sbaiz, and Martin Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct. 2006.
- [16] H.-D. Lüke, "Sequences and arrays with perfect periodic correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 3, pp. 287–294, 1988.
- [17] P. A. Naylor, J. Cui, and M. Brookes, "Adaptive algorithms for sparse echo cancellation," *Signal Processing*, vol. 86, no. 6, pp. 1182–1192, 2006.
- [18] Emanuël A. P. Habets, "Room impulse response generator," [http://home.tiscali.nl/ehabets/rir\\_generator.html](http://home.tiscali.nl/ehabets/rir_generator.html), Sept. 2010.