Sound Focusing in Rooms Using a Source with Controllable Directivity

Markus Zaunschirm, Matthias Frank
Institute of Electronic Music and Acoustics, 8010 Graz, Austria, Email: zaunschirm@iem.at

Introduction

Controlling sound in a desired region of a room is directly linked to the capability of a system to focus sound energy in time and space. Generally, hardware needed to control the soundfield consists of multiple loudspeakers and microphones. The loudspeakers can be arranged to form a loudspeaker array or used as individual loudspeakers and the microphones are required to gather information of the room characteristics. In [1] approaches for generating acoustically bright zones (higher sound energy than in other zones) and for maximizing the contrast between a bright and dark zone (no sound energy desired) are presented. A similar approach is presented in [2], where the sound energy difference is maximized. Other strategies are known from traditional cross-talk cancellation or dereverberation, where the propagation in a room between a set of loudspeakers and a set of control points is described by a multi-dimensional filter matrix. The inverse of this filter matrix can then be used to find optimal driving signals for the loudspeakers [3]. While initially used to control the pressure at discrete points, the approach can be used to control larger areas if the grid of control points is sufficiently dense. In literature, this approach is also termed as pressure matching and the minimization of error energy between a desired and produced soundfield is constrained to the sound energy in dark zones and the array effort [4]. While the mentioned methods are used for controlling larger regions of a soundfield with multiple loudspeakers or arrays, this work deals with the ability of a single loudspeaker with controllable directivity for point-to-point focusing.

Strategies for such point-to-point focusing include the time reversing of measured propagation paths [5] or obtaining an inverse of a multi-dimensional filter [6]. In order to focus the sound energy, both strategies make use of multiple propagation paths (direct sound and reflections).

Initially, the time-reversal mirror (TRM) was designed for medical applications using ultrasonic waves, and was later adapted in the field of audible range acoustics. Focusing experiments in rooms were done with a simple linear transducer array and a pressure microphone at the focal point [7] [8]. In contrast to the inverse filtering approach, the TRM focuses sound at the focal point without controlling other areas in the soundfield. However, the time reversal approach remains an attractive approach because of its simplicity. Thus, this article deals with the evaluation of the focusing quality that can be achieved using the TRM approach and a source of controllable directivity.

The theoretical concept is explained in the following section. Simulations of the concept are done for a simple shoebox room and room impulse responses are calculated using a straight forward image source method of order $N_i = 3$. Listening experiments are conducted in an anechoic chamber using a ring of 32 loudspeakers for auralization of the focused soundfield.

Theoretical Concept

Let us define an idealized, sparse response from a source with controllable directivity to the focal point with index $\hat{m}$ in a room with plane walls and without diffraction for the n^{th} time index as a superposition of individual reflection paths

$$h_{\hat{m}}[n] = \sum_l \frac{a_l}{r_l} \delta_{n,n_l}, \tag{1}$$

where a_l and r_l are the amplitude and path length. Further, $\delta_{n,n}$ is the Kronecker delta, $n_l = int\{\tau_l/\mathrm{T}\}$ is obtained by an integer rounding operation, $\mathrm{T} = 1/(44.1\ \mathrm{kHz})$ is the sampling interval and $\tau_l = r_l/c$ is the propagation delay of the l^{th} path.

With N as the length of the room impulse response (RIR) $h_{\hat{m}}[n]$, and according to [5], the TRM filter that achieves a constructive superposition of all propagation paths at the focal point is defined as

$$\hat{h}_{\hat{m}}[n] = \sum_l \frac{a_l}{r_l} \delta_{n,N-n_l} = \sum_l \hat{h}_{\hat{m},l}[n]. \tag{2}$$

With a simple impulse $\delta_{0,n}$ as source input, the signal obtained at the focal point is

$$y_{\hat{m}}[n] = \delta_{0,n} * \hat{h}_{\hat{m}}[n] * h_{\hat{m}}[n] = x[n] * h_{\hat{m}}[n], \tag{3}$$

where $*$ is the linear convolution operator and $x[n]$ is the signal emitted by the source. The concept is depicted in fig. 1.

The signal received at the focal point will have its maximum amplitude at the time instant of constructive superposition, however, additional signal components arrive before and after that. These components are termed pre- and post-echos, respectively (see fig. 1).

If additionally to the amplitude and propagation delay, the direction $\boldsymbol{\theta}_l$ of the l^{th} path is known (measured with a calibrated microphone and loudspeaker array [9]), eq. 3 can be rewritten as

$$h_{\hat{m},\hat{l}}[n] = \sum_l \frac{a_l}{r_l} g(\boldsymbol{\theta}_{\hat{l}}, \boldsymbol{\theta}_l) \delta_{n,n_l}, \tag{4}$$

where $\boldsymbol{\theta}_{\hat{l}}$ and $\boldsymbol{\theta}_l$ are Cartesian unit vectors indicating the look-direction of the source and the direction of the l^{th} reflection path, respectively.

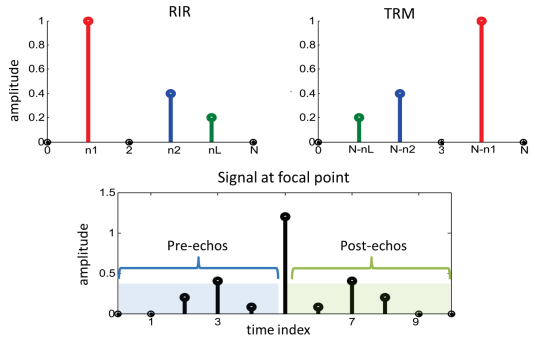


Figure 1: Concept of sound focusing with the TRM approach, an omnidirectional source and $\delta_{n,0}$ as input signal.

Modelling Source Directivity

The directivity of the source can be defined as a weighted superposition of spherical harmonics

$$g(\boldsymbol{\theta}_{\hat{l}}, \boldsymbol{\theta}_l) = \sum_{n=0}^{N_d} \sum_{m=-n}^{m=n} Y_n^m(\boldsymbol{\theta}_{\hat{l}}) w_n Y_n^m(\boldsymbol{\theta}_l), \qquad (5)$$

where $Y_n^m(\boldsymbol{\theta})$ are spherical harmonics evaluated at the direction $\boldsymbol{\theta}$, N_d is the order of the directivity pattern and w_n are normalized weights [10]. Source directivity patterns for $N_d = 3$ and different weightings w_n are shown exemplarily in fig. 2.

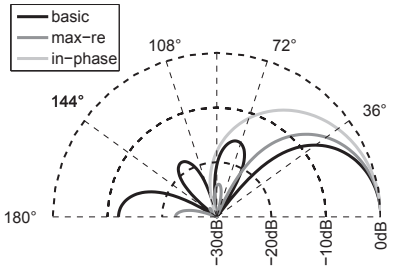


Figure 2: Third-order directivity pattern with *basic* (black), *max-re* (gray), *in-phase* (silver) weights w_n. Directivity patters are normalized such that $g(\boldsymbol{\theta}, \boldsymbol{\theta}) = 1$.

When steering the source towards each propagation path the signal at the focus is obtained by

$$y_{\hat{m}}[n] = \sum_{\hat{l}=1}^{L} \hat{h}_{\hat{m},\hat{l}}[n] * h_{\hat{m},\hat{l}}[n], \qquad (6)$$

where $\hat{h}_{\hat{l},\hat{m}}[n] = (a_{\hat{l}}/r_{\hat{l}})\ \delta_{n,N-n_{\hat{l}}}$. In words, the source input signal is weighted and delayed according to the filters defined by the TRM approach and then played back with a directivity pattern that is steered towards the corresponding propagation path. Paths different from $\hat{l} \neq l$ are damped according to eq. 5 and thus, the amount of pre- and post-echos can be reduced (compare fig. 3 and fig. 1) depending on the chosen directivity order N_d (for sufficiently high orders pre- and post-echos are entirely cancelled).

Furthermore, the sound energy emitted by the source is reduced according to the directivity index of the chosen directivity pattern and thus, the spatial focusing is assumed to be improved as well.

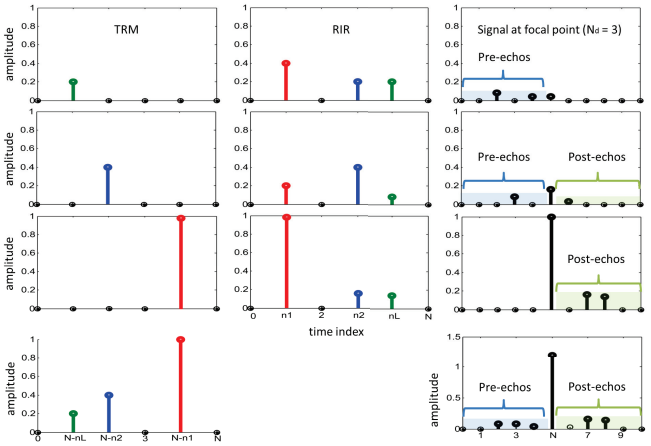


Figure 3: Concept of TRM using a source with controllable directivity.

Amplitude Compensation

As the time reversal operation does not compensate for attenuations during propagation (e.g. reflections at not perfectly rigid walls) the attenuations are squared by re-emitting the time-reversed RIR (signal obtained at focal point is filtered with the squared amplitude spectrum of the RIR). Moreover, the direct path ($l = 1$) contributes strongest to the overall energy at the focal point. In order to attain equal amplitude for all paths, they can be weighted by

$$\beta_{\hat{l}} = \frac{r_{\hat{l}}}{a_{\hat{l}}}. \qquad (7)$$

However, the amplification of paths should not increase the ratio of energy contained in pre-echos to the energy obtained at the time of coherent superposition as we assume that this would degrade the perceived audio quality. Thus, it is desired to find a compensation scheme that allows for reducing pre-echo energy. A straight forward approach is to only amplify paths with index $\hat{l}$ where $(a_{\hat{l}}/r_{\hat{l}}) > (a_l/r_l) g(\boldsymbol{\theta}_{\hat{l}}, \boldsymbol{\theta}_l)$ for $1 \leq l < \hat{l} - 1$ (see second row of fig. 3). A more elaborate weighting scheme is subject to future research.

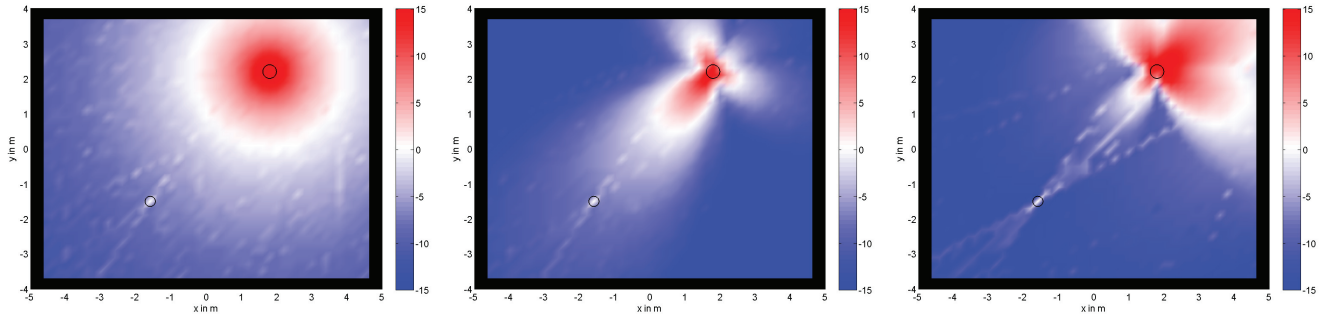
Simulations

First simulations are done for an empty room with dimensions 10×8 m^2, but can be easily extended to a room in three dimensions.

Setup

The chosen setup is depicted in fig. 6. The source with controllable directivity of order $0 \leq N_d \leq 3$ is positioned at $(1.8, 2.2)$ m, with the origin of the coordinate system in the middle of the room. The focal point is at $(-1.6, -1.5)$ m and control points are equally spaced in the room with a 20 cm spacing (overall 1900 points are evaluated).

The sparse RIR between the source position and the focus point with index $\hat{m}$ is calculated using a two-dimensional image source method of order $N_i = 3$ (25 distinct propagation paths). The walls are simulated with a frequency-



(a) $N_d = 0$ (b) $N_d = 3$ (c) $N_d = 3$ and amplitude compensated

Figure 4: Spatial focusing in dB. Black circles indicate the source and focal point, respectively.

independent reflection coefficient of 0.8 (common for rooms with a volume between $100 \ldots 5000$ m^3 with reverberation times between $0.4 \ldots 2$ s.)

As input signal an impulse $\delta_{n,0}$ is used and the source is simulated as having a usable frequency range between 300 Hz $\ldots$ 6 kHz and frequency-independent directivity within that range. The order weightings w_n are chosen for each steering angle $\boldsymbol{\theta}_{\hat{l}}$ such that the radiated energy for that path is minimized

$$\arg\min_{w_n} h^2_{\hat{m},\hat{l}}[n]. \tag{8}$$

Spatial Focusing

We compute the spatial focusing for the m^{th} point in the control region as

$$E_m = \frac{\sum_n y^2_m[n]}{\sum_n y^2_{\hat{m}}[n]}, \tag{9}$$

where $y_m[n]$ is the received signal.

It can be seen from fig. 4, that the spatial focusing is clearly improved when using a source with controllable directivity ($N_d = 3$) instead of an omnidirectional one. Although the obtained energy distribution depicted in fig. 4(b) is similar to a 3$^{\text{rd}}$ order directivity pattern, the energy at the focal point is approx. 6dB higher than at adjacent points. If amplitude compensation is used, paths $\boldsymbol{\theta}_{\hat{l}}$ where the direct path is strongly attenuated are played back with higher gain. This yields higher energy close to the source in the upper right corner of the simulated room, but an improved spatial focusing around the focal point (see fig. 4(c)).

Temporal Focusing

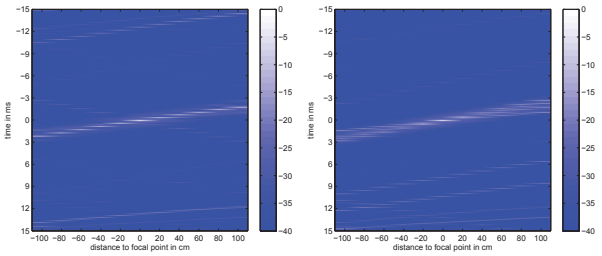
Ideally, the entire energy of the signal received at the focal point is concentrated in a small time window around the peak with time index N. We use a window of 0.45 ms (20 samples at $f_s = 44.1$ kHz) around the temporal peak to calculate the temporal focus as

$$\Gamma_{\hat{m}} = \frac{\sum_{n=N-10}^{N+10} y^2_{\hat{m}}[n]}{\sum_n y^2_{\hat{m}}[n]}. \tag{10}$$

It can be seen in tab. 1, that using a source with controllable directivity can improve temporal focusing and thus, temporal spreading is reduced. While for $N_d = 0$ only 56% of the signal energy is concentrated in a 0.45 ms time window around the time instant of coherent superposition, we obtain 86% of the energy in the same time window for $N_d = 3$. When compensating for attenuations of paths during propagation (see. eq. 7) the amount of pre-echos can be further reduced (see fig. 5)

	Γ	Γ_{pre}	Γ_{post}
$N_d = 0$	56%	22%	22%
$N_d = 3$	86%	7%	7%
$N_d = 3[comp.]$	87%	3%	10%

Table 1: Temporal focusing of the transmitted signal at the focal point. Γ_{pre} and Γ_{post} corresponds to the energy contained in pre- and post-echos, respectively.



(a) $N_d = 3$ (b) $N_d = 3$ and amplitude compensated

Figure 5: Spatio-temporal focusing in dB for control points along the x-axis with focal point in center.

Preliminary Listening Experiment

The auralization of the simulated focusing techniques is done in an anechoic chamber. A ring of 32 Genelec 8020 loudspeakers (11.25° inter-loudspeaker spacing) with a radius of 1.5 m was placed at ear height of a dummy head (1.25 m) positioned in the center of the ring.

With $\theta_{R,l}$ as the ideal incidence angle at the focus point (corresponding to l^{th} path), the loudspeaker playing back the corresponding signal is defined as

$$r = int\{\theta_{R,l}/\mathrm{A}\}, \tag{11}$$

where $\mathrm{A} = 360°/32$.

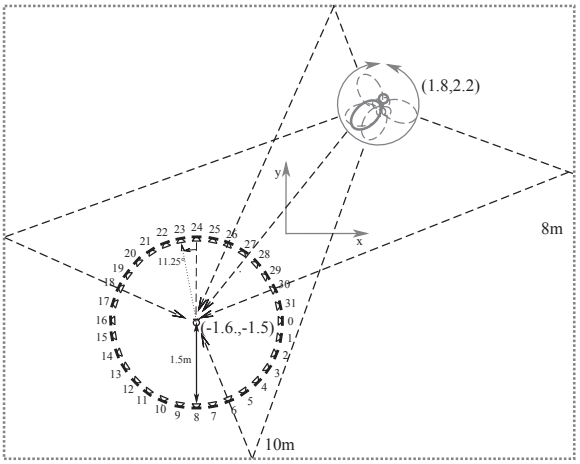


Figure 6: Simulation and auralization setup.

As test signal we used a female speaker and recorded the played back signal with one microphone, two microphones in ear distance (approx. 15 cm), and a dummy head in the center of the loudspeaker ring. In a preliminary listening experiment subjects were asked to rate the audio quality of the recorded signal compared to the speech stimulus convolved with the RIR for $N_d = 0$ and $N_d = 3$ steered towards the focal point. The simulated signal (refers to the signal obtained by convolving the test stimulus with the idealized TRM filter according to eq. 3 in Matlab) was included for verifying the playback system. First results show a clear improvement of perceived audio quality for sound focusing using a source with controllable directivity. The ratings of the binaural recordings (dummy head and spaced microphones) indicate higher audio quality of the amplitude compensated approach (abbreviated as [ac]). However, more detailed evaluations need to be done.

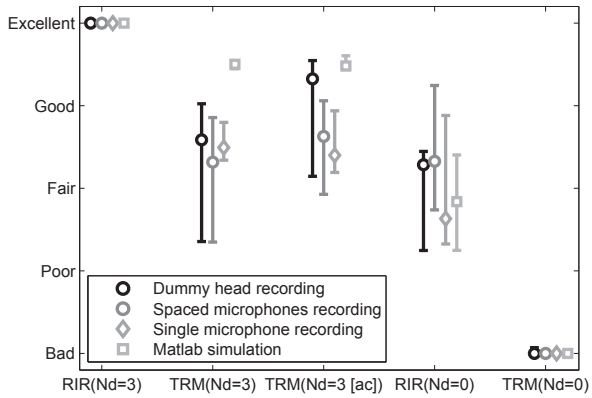


Figure 7: Perceived audio quality for different settings with a female speaker sample as stimulus.

Conclusion and Future Work

We showed that a source with controllable directivity can be used to improve point-to-point focusing in a room by individually exciting and weighting of propagation paths. Both, spatial and temporal focusing as well as the perceived audio quality of the focused signal are improved when compared to using an omnidirectional source.

However, practical directional sources (e.g. spherical loudspeaker arrays with suitable signal processing) do not generate frequency-independent directivity patterns (order N_d decreases as frequency decreases) and thus, limitations as outlined in [11] need to be included in simulations. Moreover, obtaining the direction of different propagation paths $\boldsymbol{\theta}_l$ is not a trivial task in practice and needs further investigation.

Acknowledgment

This work was partly supported by the project ASD, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions.

References

[1] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, p. 1695, 2002.

[2] M. Shin, S. Q. Lee, F. M. Fazi, P. a. Nelson, D. Kim, S. Wang, K. H. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces.," *The Journal of the Acoustical Society of America*, vol. 128, pp. 121–31, 2010.

[3] O. Kirkeby and P. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, 1999.

[4] T. Betlehem and P. D. Teal, "A Constrained Optimization Approach for Multi-Zone Surround Sound," *Ieee Icassp*, vol. 1, pp. 437–440, 2011.

[5] M. Fink, "Time reversal of ultrasonic fields. I. Basic principles," *IEEE Transactions on Ultrasonics Ferroelectrics, and Frequency Control*, vol. 39, no. 5, 1992.

[6] M. Tanter, J. L. Thomas, and M. Fink, "Time reversal and the inverse filter.," *The Journal of the Acoustical Society of America*, vol. 108, pp. 223–34, 2000.

[7] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," *The Journal of the Acoustical Society of America*, vol. 113, no. 3, p. 1533, 2003.

[8] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms. II. The spatio-temporal inverse filter," *The Journal of the Acoustical Society of America*, vol. 114, no. 6, p. 3044, 2003.

[9] H. Morgenstern, F. Zotter, and B. Rafaely, "Joint spherical beam forming for directional analysis of reflections in rooms," *The Journal of the Acoustical Society of America*, vol. 131, pp. 3207–3207, 2012.

[10] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia.* PhD thesis, 2001.

[11] H. Pomberger, *Angular and radial directivity control for spherical loudspeaker arrays.* Master thesis, 2008.