# Sound Field Control Using a Limited Number of Loudspeakers

**Conference Paper** · June 2009

CITATIONS
8

READS
174

**4 authors**, including:

Mathias Johansson
Dirac Research AB
**17** PUBLICATIONS **168** CITATIONS
SEE PROFILE

Lars-Johan Brannmark
Dirac Research AB
**20** PUBLICATIONS **111** CITATIONS
SEE PROFILE

Mikael Sternad
Uppsala University
**189** PUBLICATIONS **3,504** CITATIONS
SEE PROFILE

# Sound Field Control Using a Limited Number of Loudspeakers

Mathias Johansson[1], Lars-Johan Brännmark[1], Adrian Bahne[1], and Mikael Sternad[1]

[1]*Dirac Research AB, Uppsala, SE-75450, Sweden*

Correspondence should be addressed to Mathias Johansson (mj@dirac.se)

**ABSTRACT**

We treat the problem of optimizing $N$ digital filters for a sound system consisting of $N$ loudspeaker inputs and $M$ measurement output positions, with the objective of creating a target sound field described by $M$ prescribed transfer functions, one for each measurement position. We pose the problem in a control-theoretic framework, with the main interest on practical car audio systems, i.e. where $N \ll M$.

The solution is verified in listening experiments and by objective means. It is shown that the proposed framework provides a unified solution to the problem of equalizer design, crossover design, delay and level calibration, sum-response optimization, and up-mixing (the process of routing 2-channel or 5.1-channel source material to $N$ loudspeaker outputs).

## 1. INTRODUCTION

The general tuning process of car audio systems today proceeds in several steps. First, crossover filters are set and each loudspeaker is equalized on a per-channel basis; then the delay and level for each channel is set to reach a desired sound stage; additional adjustments to filter responses are made with respect to the combined acoustic loudspeaker responses; finally, parameters for up-mixing are adjusted. Up-mixing here refers to the process of distributing stereo or discrete 5.1 material to the $N$ loudspeakers in the car. The up-mixing problem has generated several adaptive algorithms that in realtime aim at separating 2-channel recorded sound into parts that correspond to the typical 5.1 loudspeaker setup. In addition, this has to be up-mixed (or down-mixed) further to the physical loudspeaker setup that makes up the car audio system [1, 2].

The end goal of the car tuning process can be described in terms of a *target sound field* inside the car. The target sound field is in general continuous in space, but a key insight is that a solution can be found if we alleviate the requirement on perfect reconstruction of the target sound field and further limit our target to cover only a finite number of measurement positions. By sampling the sound field at a limited number $M$ of positions inside the car, positions that with adequate resolution cover all relevant listener positions, we discretize the problem and can work directly with $N \times M$ transfer functions.

The authors have worked for several years on a control-theoretic formulation that addresses the general car tuning problem in a jointly optimal way. We here report on our findings, which include a general algorithmic framework, a number of practical implementations and reports from listening experiments. Although we here focus on the general car audio tuning problem, the same framework is applicable without modification to other applications where a target sound field is to be achieved with a limited number of imperfect loudspeakers.

### 1.1. Relations to Previous Work

In the present work, we are concerned with the reconstruction in time and space of a particular desired sound field. In the literature, there are essentially three different theoretical approaches to this problem:

1. Wave Field Synthesis (WFS), which is based on Huygens Principle, or the Kirchhoff-Helmholtz integral representation of sound fields [3].

2. High Order Ambisonics (HOA), based on a Fourier-Bessel series expansion of the original and desired sound fields in spherical coordinates [4].

3. Multipoint Mean Square Error (MSE) based methods, in which the error between the desired and the reconstructed sound field is minimized on a discrete grid of measurement points [5].

It has been shown [4, 6] that all of the three approaches above are equivalent in the limiting case when the number of transducers (loudspeakers and microphones) tend to infinity. (For the HOA to be equivalent to the WFS and MSE methods, there is a further requirement that a spherical microphone array is used.) In the present work however, a basic assumption is that the number of loudspeakers is low. Under these and other common practical circumstances, the direct multipoint MSE method is to be preferred, for reasons summarized below:

1. WFS is based on ideal assumptions regarding the transducers and the acoustic environment where the reproduction takes place. In practical systems, these assumptions are never fulfilled, and it is now widely accepted that WFS systems need a complementary measurement-based multichannel equalization (using e.g., some MSE method) to be fully functioning [7, 8].

2. The HOA approach aims at sound field reconstruction within one single spherical region, and is thus not suitable for reproduction over arbitrary spatial regions. In a car for example, we are interested in controlling the sound field mainly within four spatially disjoint regions, located around the passengers' heads.

3. If the number of transducers is low and the acoustical environment is complex, as is the case in a car, a reconstruction error is inevitable. The general multipoint MSE solution, if posed properly, offers explicit means to control the time-, space- and frequency aspects of this error.

4. The Fourier-Bessel representation in HOA is given on a per-frequency basis, and the overall filter design is performed for each frequency separately. The time domain aspect of the error, which is crucial for the perceived sound quality, is thus overlooked.

Given that the target sound field is to be obtained by minimization of the multipoint MSE, there are still many solution techniques to consider. In the present work, the MSE-optimal MIMO filters are designed using the framework of linear-quadratic feedforward control. As we shall see, this approach offers some advantages which are difficult to obtain with other methods. First, it admits explicit control over the length of the "pre-response" part of the filters. In a pure frequency domain solution, which is computationally attractive, this is normally handled by the less precise technique of regularization [9]. Second, we avoid the computational complexity normally associated with time domain matrix-based MSE methods [10], which require the inversion of extremely large block-Toeplitz matrices. Third, it imposes no restrictions on filter order or structure, as is the case in matrix-based designs, which impose a fixed-length FIR filter structure. For a given target sound field defined over the grid of measurement points, the design equations specify the unique MSE-optimal multivariable IIR compensator.

Finally we point out that the method presented here incorporates so-called transaural stereo as a special case, if the measurement points are positioned at the ear canals of one or several listeners, and target responses are properly selected [11]. A transaural setup was implemented in an automotive setting in [12]. The low number of measurement points in the transaural approach however makes it numerically ill-conditioned and spatially nonrobust.

## 2. OVERVIEW OF PROPOSED FRAMEWORK

In our present framework, the car tuning problem consists of designing $N$ digital filters, one for each loudspeaker, so that when a source (e.g., the left signal on a stereo recording) is played through these filters and the subsequent loudspeakers, a target sound field (e.g., one that would result from an ideal loudspeaker positioned at some angle and distance from the listening position in a well-designed listening room) is achieved at the $M$ measurement positions.

The overall procedure consists of the following steps:

1. Estimate the impulse responses of each loudspeaker at each measurement position.

2. Define the target stage, that is, the target impulse response at each measurement position.

3. Create an optimal controller in the form of an IIR or FIR filter matrix that makes the matrix sum-response as similar as possible to the target impulse response matrix.

4. Postprocess the optimal controller to reach a target overall tonal balance (i.e., magnitude response).

5. Subjective evaluation, with subsequent target stage adjustments and target magnitude response adjustments based on listener feedback.

The following sections describe each step in more detail.

## 3. ESTIMATION OF IMPULSE RESPONSES

The room-acoustic impulse responses of each loudspeaker at each listener position are estimated from measurements at $M$ positions where we have used $M = 64$ for the designs presented here. In these designs, the car is split into four listener seats (front left, front right, rear left, rear right). At each seat, a quadratic horizontal grid of $4\times4$ measurement positions is employed.

## 4. TARGET STAGE DEFINITION

The target stage is composed of $M$ desired impulse responses (or equivalently, transfer functions), one for each of the 64 measurement positions described in the previous section. One target stage is defined for each of $L$ virtual sources that are to be created. For example, in the case of reproducing stereo material via two virtual loudspeakers, two target stages are defined.

The target stage can be measured inside a reference listening room using the same technique as when measuring impulse responses in the car, or the target stage can be simulated. We have mainly investigated the latter approach as this simplifies experimenting and making minor changes to the target stage.

For experimental purposes, a simulator was built that allows the user to specify the target room size and loudspeaker positions within the room. Based on this input, a plane wave simulation creates $M = 64$ target impulse responses according to the measurement positions. The target impulse responses are currently limited to the direct wave and the first order reflections. The direct wave and each reflection can be assigned individual magnitude response characteristics. Various types of diffusion can be added to each wall reflection, and individual reflection levels and delays can be adjusted manually. This way, a target stage can be defined that satisfies psychoacoustic preferences pertaining to reflections in rooms, rather than being limited by physical constraints encountered in real listening rooms.

## 5. OPTIMAL CONTROLLER DESIGN

A set of $N$ filters are to be designed for each of $L$ source signals $r_\ell(t)$, $\ell = 1,\dots L$, with the purpose of minimizing the sound field approximation error at the $M$ measurement positions in some well-defined sense. We will define and use a multipoint MSE criterion that involves a possibly frequency-weighted sum of the powers of the approximation errors at all the measurement positions.

We here assume all the involved dynamic systems to be linear and time-invariant. The processing is performed in discrete time (indexed by $t$) at an appropriate sampling rate, here normalized to one. The resulting filters should be stable and also causal; they should operate only on past and present samples of the source signal.

The problem can then be solved by using known results from control theory: It becomes a special case of a Linear Quadratic Gaussian optimal (LQG) control problem, see e.g. [13], namely the design of a LQG-optimal feedforward controller. This filter design could be cast in state-space form, but we will here present and discuss a transfer function based solution in polynomial form, that offers more physical insight [14]. The solution is outlined below, after first introducing the required notation.

### 5.1. Background and Notation

The trace and transpose of a matrix $\mathbf{M}$ will be denoted $\mathrm{tr}\mathbf{M}$ and $\mathbf{M}'$, respectively. Causal FIR filters with real valued coefficients $\{p_n\}$ are here represented by polynomials in a backward time shift operator $q^{-1}$

$$P(q^{-1}) = p_0 + p_1 q^{-1} + \dots + p_{np} q^{-np} \quad .$$

The backward shift operator performs a time shift $v(t-1) = q^{-1}v(t)$ of signals , so $P(q^{-1})v(t) = p_0 v(t) + p_1 v(t-1) + \dots + p_{np} v(t-np)$. The operator $q^{-1}$ corresponds to $z^{-1}$ or $e^{-j\omega}$ in the frequency domain. A forward shift operator $q$ performs $v(t+1) = qv(t)$ and corresponds to $z$ or $e^{j\omega}$. Matrices with polynomials (FIR filters) as elements are denoted *polynomial matrices*, here represented by bold italics, $\boldsymbol{P}(q^{-1})$.

Linear time-invariant multiple-input multiple-output (MIMO) dynamic systems with finite state space can be represented by matrices with rational transfer functions as elements. These *rational matrices* are here represented by calligraphic symbols such as $\mathscr{R}(q^{-1})$. Arguments ($q^{-1}, q, z^{-1}$ or $z$) of polynomial or rational matrices are omitted where there is no risk of misunderstanding.

A rational matrix may be parameterized in terms of polynomial matrices as a matrix fraction description (MFD), either left MFD $\mathscr{R} = \boldsymbol{A}_1^{-1}\boldsymbol{B}_1$ or right MFD $\mathscr{R} = \boldsymbol{B}_2\boldsymbol{A}_2^{-1}$ [15]. A dynamic system represented by a rational matrix is stable if all its transfer function elements are stable. It is causal if all its elements are causal operators. A discrete time dynamic system represented by the inverse $\boldsymbol{P}^{-1}(q^{-1})$ of a square polynomial matrix $\boldsymbol{P}(q^{-1})$ will be

asymptotically stable if all zeros of $\det[\boldsymbol{P}(z^{-1})]$ are located in $|z| < 1$. $\boldsymbol{P}(q^{-1})$ is then sometimes called a "stable" polynomial matrix.

For a causal polynomial matrix $\boldsymbol{P}(q^{-1})$ its corresponding noncausal *conjugate matrix* $\boldsymbol{P}_*(q)$ is defined by transposing the matrix and substituting the forward shift operator $q$ for $q^{-1}$ as argument in all polynomials.

### 5.2. Model Parameterizations

Using the notation introduced above, the acoustic channel from the $N$ loudspeakers to the $M$ measurement positions is represented by

$$y(t) = \mathscr{H}(q^{-1})u_\ell(t) = \boldsymbol{B}_d(q^{-1})\boldsymbol{B}(q^{-1})\boldsymbol{A}^{-1}(q^{-1})u_\ell(t) \ . \tag{1}$$

Here, the $N \times 1$ column vector $u_\ell(t)$ represents the loudspeaker input signals at time $t$, generated by the controller in response to source signal $r_\ell(t)$, while $y(t)$ is the $M \times 1$ column vector of the sound sampled at the measurement positions. A diagonal $M \times M$ delay matrix $\boldsymbol{B}_d(q^{-1}) = \mathrm{diag}(q^{-k_i})$ contains the transport delays to each measurement position from the closest loudspeaker. The $M \times N$ polynomial matrix $\boldsymbol{B}(q^{-1})$ and the $N \times N$ stable polynomial matrix $\boldsymbol{A}(q^{-1})$ are a right MFD representation of the remaining dynamics. Only $\boldsymbol{B}(q^{-1})$ is required if a FIR model is used, in that case $\boldsymbol{A}(q^{-1}) = \mathbf{I}$.

The desired target stage for the source signal $r_\ell(t)$ is represented by vector of FIR filters

$$z(t) = \boldsymbol{U}_\ell(q^{-1})\boldsymbol{D}_\ell(q^{-1})r_\ell(t) \ . \tag{2}$$

A diagonal $M \times M$ delay matrix $\boldsymbol{U}_\ell(q^{-1}) = \mathrm{diag}(q^{-g_{\ell i}-d})$ here contains the desired delay $g_{\ell i}$ for source $\ell$ at at measurement position $i$, plus an optional added bulk delay $d$ that is common for all positions. By allowing such a bulk delay, the overall approximation fidelity can be improved. For realizability, a desired delay at a position $i$ should of course not be set smaller than the lowest physical delay, $g_{\ell i} + d \geq k_i$.

The $M \times 1$ polynomial vector $\boldsymbol{D}_\ell(q^{-1})$ describes all amplitudes and also the relative timings of secondary reflections in the desired room response as FIR filter coefficients. The relevant average spectral properties of the source signal $r_\ell(t)$ are represented by a scalar and stable autregressive model

$$H_\ell(q^{-1})r_\ell(t) = e(t) \ , \tag{3}$$

where $e(t)$ is a white discrete-time sequence.

The target stage models (2) will be different for the $L$ virtual source signals, while the room model (1) remains the same for all virtual sources.

### 5.3. The Design Problem

The $L$ source signals $r_\ell$ are modeled as mutually uncorrelated noise processes. Under that assumption, the LQG feedforward controller design problem separates into $L$ separate problems of designing stable and causal filters with single input and $N$ outputs

$$u_\ell(t) = \mathscr{R}_\ell(q^{-1})r_\ell(t) \ , \ \ell = 1 \ldots L \ . \tag{4}$$

Define the error vector obtained by (1),(2) and (4):

$$\varepsilon(t) = z(t) - y(t) = (\boldsymbol{U}_\ell\boldsymbol{D}_\ell - \mathscr{H}\mathscr{R}_\ell)r_\ell(t) \ .$$

The scalar design criterion to be minimized weights the power of the approximation error components $\varepsilon_i(t) = z_i(t) - y_i(t), i = 1, \ldots M$ and adds optional penalty terms on the loudspeaker output signals $u_{j\ell}(t), j = 1, \ldots N$ in the form of a scalar quadratic function of these signals

$$\begin{aligned} J &= \sum_{i=1}^{M} E|V_i(q^{-1})\,\varepsilon_i(t)|^2 + \sum_{j=1}^{N} E|W_j(q^{-1})\,u_{j\ell}(t)|^2 \\ &= E\{(\boldsymbol{V}\varepsilon(t))'\boldsymbol{V}\varepsilon(t) + (\boldsymbol{W}u_\ell(t))'\boldsymbol{W}u_\ell(t)\} \\ &= E\{\mathrm{tr}[\boldsymbol{V}\varepsilon(t)(\boldsymbol{V}\varepsilon(t))'] + \mathrm{tr}[\boldsymbol{W}u_\ell(t)(\boldsymbol{W}u_\ell(t))']\}. \end{aligned} \tag{5}$$

The expectation $E(\cdot)$ in (5) is taken with respect to the statistical properties of $r_\ell(t)$. The weighting $\boldsymbol{V}(q^{-1}) = \mathrm{diag}[V_i(q^{-1})]$ is a square diagonal polynomial matrix of full rank $M$. When no frequency weighting is used, $\boldsymbol{V}(q^{-1}) = \mathbf{I}$. The square diagonal polynomial matrix $\boldsymbol{W}(q^{-1}) = \mathrm{diag}[W_j(q^{-1})]$ can be used to focus the control energy into frequency ranges that are appropriate for particular loudspeakers. Each penalty FIR filter $W_j(q^{-1})$ is then given low gain within the operating range of loudspeaker $j$ and high gain outside of that range.

### 5.4. LQG Optimal Feedforward Controller

The filter design problem formulated above becomes a special case of a MIMO LQG feedforward regulator design problem that was stated and solved in [16]. A solution exists under the following assumption:

**Assumption 1:** There exists an $N \times N$ polynomial matrix $\beta(q^{-1})$ that satisfies the spectral factorization equation

$$\beta_*\beta = \boldsymbol{B}_*\boldsymbol{V}_*\boldsymbol{V}\boldsymbol{B} + \boldsymbol{A}_*\boldsymbol{W}_*\boldsymbol{W}\boldsymbol{A} \ , \tag{6}$$

and that is stable and has nonsingular leading coefficient matrix $\beta(0)$.

Assumption 1 implies that $\beta^{-1}(q^{-1})$ represents a stable and causal dynamic system. The right polynomial matrix spectral factor $\beta(q^{-1})$ will furthermore be unique, up to an orthogonal left factor matrix. Assumption 1 is fulfilled under mild assumptions, see [16]. For example, the use of an input penalty matrix such that $\det[\boldsymbol{W}(z^{-1})] \neq 0$ on the unit circle $|z| = 1$ will guarantee it.

Under Assumption 1, the stable and causal linear feedforward regulator (4) that minimizes the criterion (5) for a model (1), (2), (3) is given by

$$u_\ell(t) = \mathscr{R}_\ell(q^{-1})\, r_\ell(t) = \boldsymbol{A}(q^{-1})\beta^{-1}(q^{-1})\boldsymbol{Q}_\ell(q^{-1})\, r_\ell(t) \ , \tag{7}$$

where the causal $N \times 1$ polynomial vector $\boldsymbol{Q}_\ell(q^{-1})$ is, together with a noncausal $N \times 1$ polynomial vector $\boldsymbol{L}_{\ell *}(q)$, the unique solution to the linear polynomial matrix equation (Diophantine equation)

$$\boldsymbol{B}_* \boldsymbol{B}_{d*} \boldsymbol{V}_* \boldsymbol{V} \boldsymbol{U}_\ell \boldsymbol{D}_\ell = \beta_* \boldsymbol{Q}_\ell + q \boldsymbol{L}_{\ell *} H_\ell \ . \tag{8}$$

See Section 3.3 of [16] for a proof. This design performs joint equalization, crossover design, delay and level calibration to approximate the prescribed sound field response (2) according to the criterion (5).

The design equations are the polynomial matrix spectral factorization (6) and the polynomial matrix Diophantine equation (8), both of which can be efficiently solved using iterative methods, see e.g. [17] for a survey of the spectral factorization problem.

It can be noted that the matrix $\boldsymbol{A}^{-1}(q^{-1})$ in the model (1) will be canceled directly by the left factor $\boldsymbol{A}(q^{-1})$ of $\mathscr{R}_\ell(q^{-1})$. The factor $\boldsymbol{A}(q^{-1})\beta^{-1}(q^{-1})$ in (7) is an $N \times N$ rational matrix with poles in $\det[\beta(z^{-1})] = 0$. These poles are inside the unit circle, since $\beta(q^{-1})$ is stably invertible under Assumption 1. Likewise, this filter is causal, since the spectral factor has nonsingular leading coefficient matrix and is therefore causally invertible.

The factor $\boldsymbol{A}(q^{-1})\beta^{-1}(q^{-1})$ in (7) remains unchanged for all source signals. The polynomial matrix $\boldsymbol{Q}_\ell(q^{-1})$ depends on the index $\ell$ of the source signal since the target stage model $\boldsymbol{U}_\ell(q^{-1})\boldsymbol{D}_\ell(q^{-1})$ and $H_\ell(q^{-1})$ are present in the Diophantine equation (8).

Since the source signals are assumed uncorrelated, energy errors and criteria values arising from different source signals are additive. The solution for the upmixing problem for $L$ virtual sources can therefore be expressed by forming a total control signal $u(t)$ to the loudspeakers as a superposition of the optimal control signals for each source:

$$u(t) = \sum_{\ell=1}^{L} \mathscr{R}_\ell(q^{-1})\, r_\ell(t) = \boldsymbol{A}\beta^{-1}[\boldsymbol{Q}_1 \ldots \boldsymbol{Q}_L]\, r(t) \ , \tag{9}$$

where $r(t) = (r_1(t), \ldots, r_L(t))'$. We thus obtain a controller matrix (9) with stable and causal IIR filters as elements. It operates on $L$ inputs (all virtual sources) and generates the $N$ loudspeaker signals as outputs.

A final step of the design is then to tune scalar filters for each source $r_\ell(t)$ that adjust the resulting spectral magnitude response, see Section 6 below.

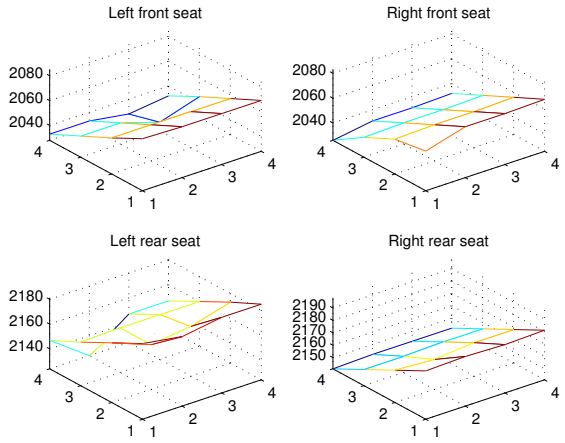


**Fig. 1:** Delay surfaces in all measurement positions for a virtual source at 0°. The horizontal plane corresponds to the 4×4 measurement positions at each seat. The vertical axis shows the measured delay in samples at 44.1 kHz sampling. The position of the vertical axis (4,1) corresponds to the front-most left-most position.

## 6. POSTPROCESSING FOR SPECTRAL SMOOTHNESS

The final step in the design of the filters is to adjust the controller response so that, on average, a target magnitude response for each virtual source is reached in the listening region. Remember that the target stage is normally designed using a simulator that creates plane wave

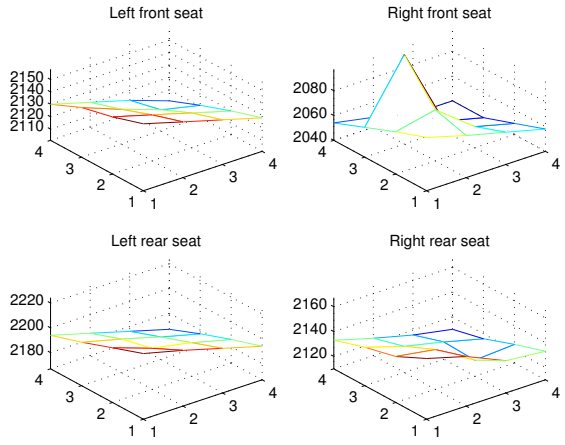


**Fig. 2:** Delay surfaces in all measurement positions for a virtual source at 45°.

impulse responses. If the target stage consists of only direct sound, the resulting frequency response is white. If in addition to the direct wave the target stage also contains reflections, spectral coloration will arise. Moreover, the designed controller matrix inevitably will have remaining approximation errors since the number of measurement positions is typically much larger than the number of loudspeakers. These errors may not be white. Magnitude response imperfections are generally undesirable and we consequently need to adjust our controller matrix so that an overall target magnitude response is reached on average in the listening region.

Hence, the magnitude responses of the overall system (including the filters) are evaluated in the various listening positions. A minimum phase filter is then designed so that on average (in the RMS sense) the target magnitude response is reached throughout the car. Variable fractional octave smoothing based on the spatial response variations is preferably employed in order not to overcompensate in any particular frequency region.

In summary, in the previous step–the design of the controller–a virtual loudspeaker in a virtual room is created which aims at replacing the physical car loudspeakers and cabin acoustics. In the postprocessing step, each virtual loudspeaker is tuned inside the virtual room to a desired tonal characteristic.

## 7. PERFORMANCE ANALYSIS

We have tested the proposed framework in a number of cars. For our evaluation here, we employed a car equipped with 7 mid-range + tweeter pairs, and 2 bass woofers (working range roughly 15–250 Hz) placed underneath the front seats. Thus, $N = 9$. The 7 "full-range" speakers have a working range of roughly 150–20 000 Hz, and were located as follows: one center speaker in the center of the dashboard, a left front and a right front speaker with the mid-range in the door, the tweeter in the A-pillar, two rear door (coaxial) speakers, and two speakers in the parcel shelf. A rather representative premium car sound system.

### 7.1. Objective Evaluation

In this evaluation we use a single plane-wave virtual source as our target stage and judge various measured performance attributes. All results were obtained using 9 FIR filters, one per physical loudspeaker, of length 8192 coefficients at 44.1 kHz sampling.

#### 7.1.1. Direction of arrival

Figures 1–3 show the measured delays at each measurement position for virtual sources placed at 0°, 45° and -120° respectively. Only minor errors occur; the delay surfaces show a clear tilt towards the intended direction. Possible sources of error include delay estimation errors and inadequate sound field reconstruction. On the whole, the algorithm reproduces sound waves from the desired directions.

#### 7.1.2. Magnitude response

Figure 4 shows the magnitude responses achieved in the front seats and rear seats with a flat target magnitude response. The virtual source was placed at 0°. Comparing the two average responses, we see that a flat mean response is achieved but that there are slight differences between the front and rear seats. Within each seat row the differences are however surprisingly small. In fact, the spatial magnitude response variations have been significantly reduced as compared to the variations of each individual physical speaker. Figure 5 shows the resulting magnitude responses from 16 different positions in the left front seat and the right front seat respectively for a virtual left front speaker. Compare this to the spatial variations from the physical left front speaker in Figure 6. The algorithm has evened out the magnitude responses over space by a proper combined use of the 9 loudspeakers.

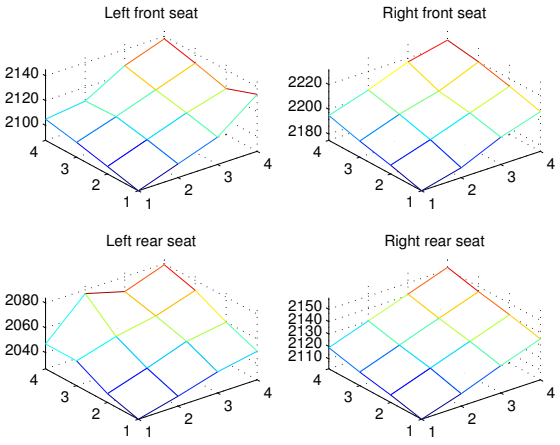


**Fig. 3:** Delay surfaces in all measurement positions for a virtual source at -120°.

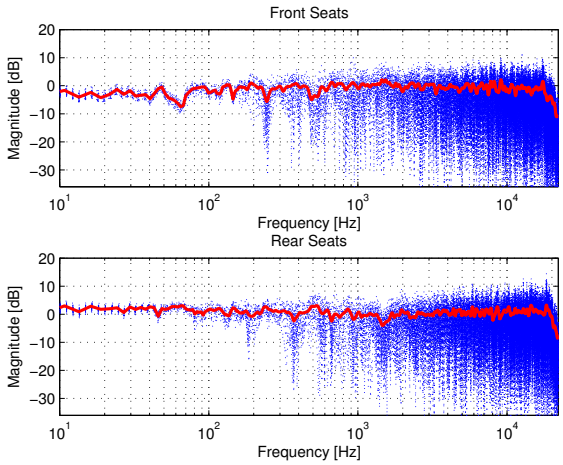


**Fig. 4:** Magnitude responses in the two seat rows for a virtual source at 0° with a flat target response.

#### 7.1.3. Time fidelity

Although we have already seen that the algorithm produces physical wave patterns with the correct direction of arrival, it is interesting to see how the impulse responses compare to the original responses and how well they resemble the optimum Dirac spike. We therefore compare the resulting impulse responses in the left front seat corresponding to the virtual source from 0° with those of the physical center speaker and the physical left front speaker (as these two speakers are closest to the intended angle of arrival and closest to the seat in question). A difficulty arises since the virtual source has a full-range response (since it uses the bass woofers) whereas the physical speakers only work down to roughly 150 Hz. As low-frequency energy typically has a longer decay, our comparison favors the physical speakers.

As a measure of *time fidelity*, or impulse response distinctness, we use the mean peak-to-tail energy in the impulse responses at the 16 microphone positions in the left front seat. We computed the ratio between the peak energy and the average energy over 1 millisecond intervals centered at $n$ milliseconds after the main peak, $n = 1...20$, in the 16 positions, then averaged over all positions. We found that the center speaker had an average peak-to-tail energy ratio of 25.2 dB, and the left front speaker 23.6 dB. The virtual 0° source reached a ratio of 26.4 dB, an improvement of 1.2 dB and 2.8 dB respectively. Averaging over a shorter time interval, up to 5 milliseconds after the main peak, the improvements become larger: 2.3 dB and 3 dB respectively. This shows that the early reflections have been reduced using the virtual speakers as compared to the two closest physical speakers. It should be mentioned that this is not true for all virtual sources and all measurement positions. On some positions, early reflections are as loud as or louder than the nearest physical speaker's original impulse response. Bear in mind however that the comparison is somewhat unfair because of the different frequency ranges. Figure 7 shows the average tail-to-peak energy ratio in the different speaker positions in the front seat for the virtual source at 0° and the left front speaker. Time fidelity has been improved using the here proposed approach.

### 7.2. Subjective Evaluation

To complete our study, we now investigate subjective aspects of the solution with regards to both individual target stage properties and the overall sound impression. We compare our results to previously published work where applicable. The stimulus used in the listening experiments is mainly speech, but often complemented by music (both 2.0 and discrete 5.1). The participating subjects represent listeners ranging from unexperienced to experienced.

#### 7.2.1. Direction of arrival

We first study the behavior of a single virtual source, e.g. a virtual left front speaker, without additional re-

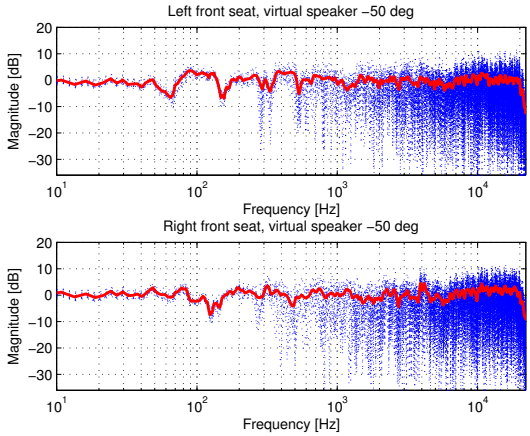


**Fig. 5:** Magnitude responses in the two front seats for a virtual source at $-50°$.

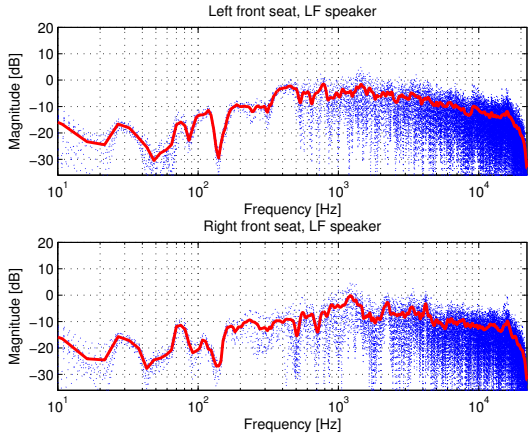


**Fig. 6:** The magnitude responses for the left front loudspeaker in the two front seats. The physical speaker has more spatial variations than the virtual speaker.

flections. We compare several angles of arrival for the virtual source (15°–75°, 5° steps) and investigate if the subjective impression of direction of origin meets the objective measurements (cf. delay surfaces, Section 7.1.1). The test showed that the subjective and objective investigations match. It is possible to distinguish between the angles, and each angle has a unique origin. Minor timbral variations between the different angles occurred.

### 7.2.2. A target stage including reflections

The next step is to add a first order lateral reflection to the target stage. It is interesting to compare our results to previous work concerning small rooms. We examine 1) the *Absolute detection threshold* (below this level, a reflection is indistinguishable from being completely absent) and 2) the *Second image audible threshold* (the reflection is perceived as an individual sound event, separated in both time and direction) in a listening experiment. The target stage corresponds to a virtual left front speaker at -30°. We investigate the left wall reflection (-70°) for delays of 5, 10 and 15 ms.

Figure 8 shows our thresholds in comparison to findings for lateral reflections from Olive and Toole [19] and Lochner and Berger [20]. The absolute detection threshold is strongly related to the test conditions and the used stimulus [18], [19], [20]. Olive and Toole's absolute detection threshold was investigated in a room with similar acoustics to the car cabin and is found to match the one examined for the present work well. It is between -10 and -15 dB for delay times up to 30 ms.

Our second image audible threshold is constant at 0 dB below 10 ms and decreases afterwards. In contrast, Lochner and Berger's threshold increases in the same range with the maximum at around 15 ms after which it decreases. This result is however not surprising considering the different experimental conditions.

In other experiments, we asked subjects to choose their "preferred" reflection level. The preferred reflection levels lie significantly ($> 8$ dB) below the second image audible threshold. It can be noted that reflections originating from a similar direction as the direct wave caused audible comb filtering. Subjects preferred the direct wave without any reflection under such circumstances.

### 7.2.3. Frequency response of reflections

We now consider the influence of reflection spectra in our solution. Previous work [19, 20, 22] shows that the frequency response of an individual reflection becomes less critical when more reflections are present. Under realistic room conditions, the influence of an individual reflection is negligible. This is in accordance with our results. Changing the target spectrum of just one reflection was hardly audible. Changing the target spectrum of all reflections together however had a significant influence on the perceived sound impression. Even changing the spectra of, e.g., all lateral reflections had an audible, albeit smaller, impact. Instead of reducing the high frequency

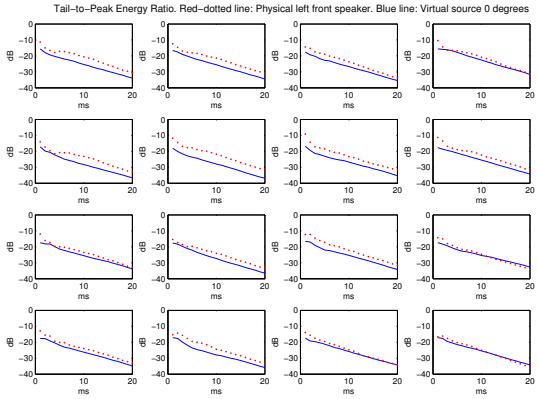


**Fig. 7:** Tail-to-peak energy ratios at the 16 measurement positions in the left front seat. The red-dotted line corresponds to the physical left front speaker, the blue line to the virtual source at 0°.

content to match a realistic frequency response, our investigation resulted in a more unusual target spectrum for reflections (Figure 9). Applying a low pass at around 1 – 2 kHz resulted in better decoupling from the physical loudspeakers and a more natural timbre of voices. A high pass at about 200 Hz was capable of reducing time-domain interference at low frequencies. The result was a cleaner bass response and a slightly better localization and image.

### 7.2.4. Overall sound impression

Finally, we investigate a complete target stage consisting of a virtual 5.1 home-theater set-up. The virtual room and loudspeaker characteristics were determined from tests similar to those reported above. Subjects experienced increased distance to the sound sources and an enlarged listening space as compared to a traditionally tuned car sound system. All subjects experienced that the sound sources were located outside the car. They further mentioned a relaxed listening impression, a pleasant wide stage, good resolution and localization of sound sources for both stereo and surround material. The statements were independent from the seating position in the vehicle.

Several listening tests examined during the present work proved the proposed concept to work properly. The solution is capable of reproducing a target stage that subjectively and objectively resembles the intended impression.

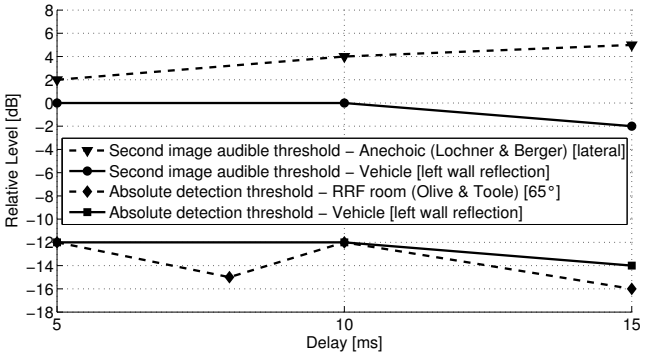


**Fig. 8:** Absolute and second image audible detection thresholds for lateral reflections in several experiments and listening environments.

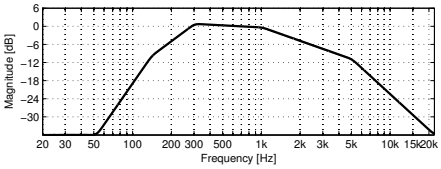


**Fig. 9:** Target spectrum of reflections in the described solution.

For further listening experiments and more detailed descriptions and results, see [21].

## 8. CONCLUSIONS

It has been verified in measurements that the proposed approach

- produces wave fronts from the desired directions,
- evens out spatial variations in magnitude responses
- improves the *time fidelity*, which we define as the mean peak-to-tail energy ratio of the impulse responses

Although an extensive listening evaluation is beyond the scope of this paper, several listening evaluations have shown that all passengers in the car experience a common natural sound impression, with a sound field perceived as decoupled from the physical speakers and emanating from outside the car. Subjects have consis-

tently expressed positive judgements regarding the staging, tonal balance, depth, and spaciousness of the resulting sound.

The proposed approach unifies and solves several problems that traditionally have been treated as separate issues: equalization, crossover design, delay and level adjustments, sum-response optimization, and up-mixing. Conversely, our approach can be viewed as an attempt to alleviate several limitations of wave field synthesis: approximate reconstruction is allowed, the sound field is spatially discretized over a region of arbitrary shape, and doing so allows us to directly manipulate the actual measured sound field in this region, without any a priori approximations or idealized assumptions.

## 9. REFERENCES


[1] H. Lahti, "Dolby surround systems in cars", *Proceedings of the AES 15th International conference on audio, acoustics and small spaces*, pp. 102–106, October 1998.

[2] T. Nind, "Multimedia in cars: The use of Logic 7 surround processing as the solution to the challenge of providing surround sound in cars from all 2 channel and encoded 5.1 sources", Presented at *AES 110th Convention, New York*, preprint 5286, April 2001.

[3] S. Spors, R. Rabenstein and J. Ahrens, "The theory of wave field synthesis revisited", Presented at *AES 124th Convention, Amsterdam*, preprint 7358, May 2008.

[4] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics", *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, November 2005.

[5] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters", *J. Audio Eng. Soc.*, vol. 37, no. 11, pp. 899–907, November 1989.

[6] F. M. Fazi, and P. A. Nelson, "A theoretical study of sound field reconstruction techniques", *Proceedings of the 19th international congress on acoustics*, Madrid, September 2007.

[7] P.-A. Gauthier and A. Berry, "Adaptive wave field synthesis for sound field reproduction: Theory, experiments and future perspectives", *J. Audio Eng. Soc.*, vol. 55, no. 12, pp. 1107–1124, December 2007.

[8] E. Corteel, "Equalization in an extended area using multichannel equalization and wave field synthesis", *J. Audio Eng. Soc.*, vol. 54, no. 12, pp. 1140–1161, December 2006.

[9] O. Kirkeby, P. A. Nelson, H. Hamada and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization", *IEEE Transactions on Speech and Audio Processing*, vol. 6 no. 2, pp. 189–194

[10] Y. Huang, J. Benesty and J. Chen, *Acoustic MIMO Signal Processing*, Springer, New York, 2006

[11] J. Bauck and D. H. Cooper, "Generalized transaural stereo and applications", *J. Audio Eng. Soc.*, vol. 44, no. 9, pp. 683–705, September 1996.

[12] Y. Kahana,, P. A. Nelson and S. Yoon, "Experiments on the synthesis of virtual acoustic sources in automotive interiors ", *Proceedings of the AES 16th International conference on Spatial Sound Reproduction*, pp. 218–232, March 1999.

[13] H. Kwakernaak and R. Sivan, *Linear Optimal Control Systems*, Wiley, New York, 1972.

[14] A. Ahlén and M. Sternad, "Wiener filter design using polynomial equations", *IEEE Transactions on Signal Processing*, vol. 39, pp. 2387–2399, 1991.

[15] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

[16] M. Sternad and A. Ahlén, LQ Controller Design and Self-tuning Control. Chapter 3 in K. Hunt Ed, *Polynomial Methods in Optimal Control and Filtering*, Control Engineering Series, Peter Peregrinus, London, Jan. 1993, pp. 56–92.

[17] V. Kŭcera, "Factorization of rational spectral matrices: a survey of methods", *IEE Control 91*, Edinburgh, pp. 1074–1078, 1991.

[18] D.R. Begault, B.U. McClain, M.R. Anderson, "Early reflection thresholds for virtual sound sources", proceedings of the *2001 international workshop on spatial media*, Aizu-Wakamatsu, Japan (2001, Oct)

[19] S.E. Olive, F.E. Toole, "The detection of reflections in typical rooms", *J. Audio Eng. Soc.*, Vol.37, No.7/8, pp.539-553 (1989, Jul/Aug)

[20] F.E. Toole, "Loudspeakers and rooms for sound reproduction: A scientific review", *J. Audio Eng. Soc.*, Vol.54, No.6, pp.451-476 (2006, Jun)

[21] A. Bahne, "Psychoacoustic aspects of reproducing acoustic room responses in cars", Diplomarbeit at the FH OOW in Emden (2008, Mar)

[22] S. Bech, "Timbral aspects of reproduced sound in small rooms. II", *J. Acoust. Soc. Am.*, Vol.99, No.6, pp.3539-3549 (1996, Jun)