Downloaded from orbit.dtu.dk on: Mar 28, 2024



DTU Library

On the influence of transfer function noise on sound zone control in a room

Moller, Martin Bo; Nielsen, Jesper Kjaer; Fernandez Grande, Efren; Olesen, Soren Krarup

Published in:
IEEE/ACM Transactions on Audio, Speech, and Language Processing

Link to article, DOI:
10.1109/TASLP.2019.2921151

Publication date:
2019

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
Moller, M. B., Nielsen, J. K., Fernandez Grande, E., & Olesen, S. K. (2019). On the influence of transfer function noise on sound zone control in a room. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*(9), 1405-1418. Article 8732395. https://doi.org/10.1109/TASLP.2019.2921151

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

On the Influence of Transfer Function Noise on Sound Zone Control in a Room

Martin Bo Møller, *Student Member, IEEE,* Jesper Kjær Nielsen, *Member, IEEE,*
Efren Fernandez-Grande, and Søren Krarup Olesen,

***Abstract*—Sound zones are valuable in scenarios where multiple people are present in the same room but want to listen to individual audio content without wearing headphones. The purpose of sound zone methods is to minimize the acoustic leakage between the zones by controlling multiple loudspeakers. This requires knowledge of how the loudspeakers interact with the room and radiate sound to the zones. That interaction is characterized by the transfer functions between the loudspeakers and microphones sampling the sound field in the zones. In this paper, the effect on the acoustic separation due to inherent noise in in-situ transfer function measurements is investigated. The attainable separation is analyzed in the frequency range 20-300 Hz by means of the eigenfunctions of a rectangular room. The concept of observable degrees of freedom is introduced to indicate the number of active eigenfunctions which are different within the zones at a given frequency. Likewise, controllable degrees of freedom indicate whether each source can excite the active eigenfunctions independently. It is argued that high separation can be achieved when the observable degrees of freedom are fewer than the controllable, and the target sound field can be described by the observable degrees of freedom. However, to attain this high separation it is a requirement that the details in the transfer functions associated with these degrees of freedom can be resolved in the presence of the measurement noise. For both simulated and experimental conditions the transfer functions are estimated using Bayesian inference and the uncertainty in the estimates is used to automatically regularize the sound field control. This regularization is seen to improve the performance when the measurement noise is correlated between the microphones and have little effect when the noise is uncorrelated.**

***Index Terms*—Sound zones, personal audio, sound field control.**

I. Introduction

IMAGINE a group of people sitting in a room, each listening to individual audio content without disturbing each other. The occupants of the room are not wearing headphones and are able to freely communicate with one another. This is the goal of sound zones systems [1], [2], where loudspeakers distributed throughout the room are used to control sound reproduction to the occupants. Limiting the spatial extent where audio is reproduced is desirable in various applications and has been investigated in environments such as: aircraft seats [3], automotive cabins [4]–[7], domestic rooms [8]–[11], and open air concerts [12].

The concept behind sound zones is to divide the space into three regions: a region where the desired sound should be reproduced, a region where it should be suppressed, and the remainder of the environment where the sound is not explicitly controlled. The region where the sound is desired is referred to as the bright zone, and the region where it should be suppressed is referred to as the dark zone [13]. Reproducing individual audio for two listeners (content A and B) is achieved by superimposing two sound zones solutions, where the bright zone with respect to audio content A is the dark zone with respect to content B and vice versa. Perceptual studies indicate that 25 to 29 dB loudness difference[1] between the target and interfering audio is sufficient for not being distracted by the interfering audio [14].

In this paper, the frequency range of interest is 20-300 Hz, where sound can be reduced in the dark zone through destructive interference [3], [13], [15]. This low frequency control should be considered as part of a composite solution covering the audible frequency range. Higher frequencies could for instance be reproduced using beamforming [16], [17] as suggested in [1]. Controlling the sound field at low frequencies in a room requires knowledge of the transfer functions between loudspeakers and zones which are measured in-situ. The question of interest is how the room response combined with the accuracy of the measurements affect the achievable acoustic separation. Two categories of sound zone methods are investigated here: acoustic contrast control (ACC) and pressure matching (PM). ACC maximizes the ratio of squared sound pressures in the bright and dark zone [3], [13] whereas PM minimizes the mean square error of the reproduced sound field in the zones relative to a predefined target [15], [18].

Typically, sound zones are implemented as feed-forward control systems based on measured transfer functions. From the measurements, a finite impulse response (FIR) filter is determined to prefilter the audio signal for each loudspeaker.

M. B. Møller is with the research department of Bang & Olufsen A/S, Struer, DK-7600, Denmark and the Signal and Information Processing Section, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark. (e-mail: mim@bang-olufsen.dk).

J. K. Nielsen is with the Audio Analysis Laboratory, Department of Architecture, Design and Media Technology, Aalborg University, 9220 Aalborg, Denmark (e-mail: jkn@create.aau.dk)

S. K. Olesen is with the Signal and Information Processing Section, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark e-mail: (sko@es.aau.dk).

E. Fernandez-Grande is with the Acoustic Technology group, Department of Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (e-mail: efg@elektro.dtu.dk).

Manuscript received April 19, 2005; revised August 26, 2015.

[1]Loudness levels depend on both the frequency content of the reproduced audio signals and the frequency response of the reproduction system, which in audio reproduction systems is a design choice. To evaluate the results without dependence on the input audio signals and spectral emphasis of the reproduction system, the results in this paper are reported in terms of relative pressure levels in the zones assuming the input signal to be a sum of pure tones with equal amplitude.

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

Thus, it is assumed that a loudspeaker in a room behaves as a linear time-invariant system. In reality, there is a number of violations of this assumption. For one thing, the ambient temperature might change which in turn changes the transfer functions and degrades the sound field control as seen in [7], where a temperature change between $-2°$C and $22.8°$C yielded a reduction in acoustic separation between two zones from 30 dB to 10 dB. Moreover, loudspeakers become non-linear at high excursions which reduces the separation between the zones, as seen in [19] where the separation is reduced from 41 dB to 27 dB due to changes at high excursion levels. Lastly, it is assumed that the measured transfer functions are accurate representations of the physical system during reproduction. However, as measurements are never noise-free, the control is based on estimated transfer functions which inherently include uncertainty.

To ensure robust performance, it is of interest to determine the consequences to the acoustic separation when the transfer functions deviate from the estimates. In [20], Park et al. investigated the sensitivity of the ACC method relative to the variance of perturbations in the transfer functions for a line-array in free field. It was observed that the sensitivity decreased with denser microphone grids sampling the transfer functions in space. A common approach to decrease the sensitivity to general changes in the transfer functions is to add regularization when determining the filters. The relationship between Tikhonov regularization and a change in speed of sound of 10 m/s or a loudspeaker displacement of 10 mm was investigated for the PM and ACC methods, using a circular loudspeaker array under free field conditions, in [21]. The results showed higher sensitivity to the position mismatch than the speed of sound change, and that robustness increases with the introduced regularization. However, increasing the regularization also reduces the separation between the zones, and a careful choice of the regularization parameter is required. Elliott et al. showed in [22] that the separation achieved by a line array with filters designed for free field transfer functions decreases when the transfer functions are changed due to adding reverberation from a room as a diffuse contribution. It was seen that given knowledge of the energy contained in the diffuse contribution the sensitivity to the contribution could be reduced. These results were experimentally validated in [8]. In [23]–[25], Zhu et al. investigated the robustness of sound zones using an arced loudspeaker array, in free field and in an acoustically treated recording studio, with transfer functions which are uncertain characterizations of the contribution from each loudspeaker to the sound field in the zones. It was shown that given the loudspeaker and microphone positions and a rough knowledge of the deviation from the transfer functions, it is possible to determine a regularization parameter which makes the sound field control robust to this deviation. The underlying premise in these investigations is that the transfer functions are known, and that regularization is added to compensate for the expected deviations. Furthermore, the typical environment is either acoustically treated or free field simulations and the loudspeaker arrangements are limited to line or arc arrays at mid and high frequencies. Under such conditions, the sound zone problem resembles super directive beamforming. In the present study, only the estimated transfer functions are known and the uncertainty in the estimates is used to regularize the solution. The sensitivity is investigated at low frequencies where the interaction between room and loudspeakers is poorly modeled by the free field radiation model with a diffuse sound contribution [26]. Instead, the sensitivity is investigated using Green's function in a rectangular room and experimental measurements in a room with a reverberation time (T_{20}) of 0.6 s in the frequency range of concern.

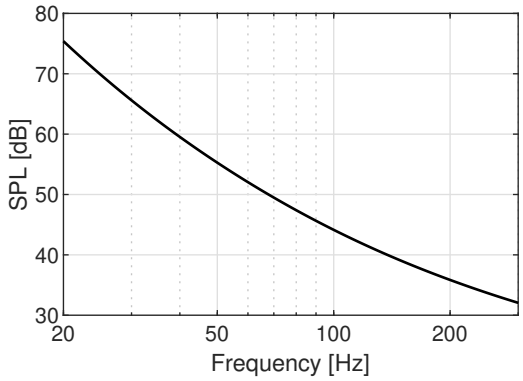


Fig. 1. Potential background noise in domestic rooms during day time due to the $L_{\mathrm{pA,LF}} \leq 25$ dB recommendation in [27].

The application assumed in this paper is a domestic room. As rooms are different and loudspeaker placement depends on interior decoration, it is necessary to measure the transfer functions in-situ. In that scenario, it is of interest to consider the amount of background noise which might be present during the transfer function measurements. In Denmark, the Environmental Protection Agency recommends A-weighted low frequency (10 - 160 Hz) noise limits of $L_{\mathrm{pA,LF}} \leq 25$ dB in domestic spaces during day time [27]. This A-weighting heavily attenuates low frequency sound, thus, significant low frequency noise can be present in a domestic setting without exceeding this recommendation as seen in Fig. 1. Assuming the maximal sound pressure level produced by the loudspeakers is limited due to non-linear distortion, the signal to noise ratio (SNR) of the transfer function measurements can be significantly reduced due to this limitation on the measurement level.

The influence of background noise during the transfer function measurements has previously been considered in [28] where an experimental study was presented. The purpose of the study was to investigate the variability in sound zone performance over repeated transfer function measurements with and without regularization based on the sample variance. In this paper, the challenges associated with the sound zones problem are related to the eigenfunction expansion of sound fields in rooms and the accuracy of the estimated transfer functions. To quantify this accuracy, marginal probability distributions for the estimates are determined using Bayesian inference. It is shown how these distributions can be used to automatically determine a regularization parameter for the PM and ACC methods under both simulated and measured conditions.

The rest of the paper is structured as follows: in section II

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

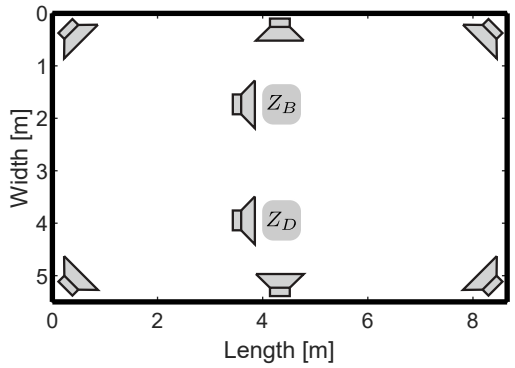


Fig. 2. Top view sketch of the layout of loudspeakers and zones in the room used for both simulations and measurements throughout the paper. Z_B denotes the bright zone and Z_D the dark zone.

the signal model and the sound zones problem are introduced. The properties of the pressure matching problem are analyzed in section III, before simulation and experimental results are introduced in sections IV and V. The results are discussed in section VI and concluding remarks are presented in section VII. The Bayesian inference procedure used to estimate the transfer functions is described in the appendix.

II. Formulating the Problem

The scenario treated in this paper is controlling the sound field in two zones: one bright and one dark as depicted in Fig. 2. The first step is to define the sound pressure observed at the m^{th} microphone at discrete time n due to the sum of filtered responses from L loudspeakers. This can be expressed as

$$\tilde{p}_m[n] = \sum_{\ell=1}^{L} \sum_{k=0}^{K-1} \sum_{i=0}^{I-1} \tilde{w}_\ell[k] \tilde{h}_{m\ell}[i] \tilde{s}[n-k-i], \tag{1}$$

where $\tilde{\cdot}$ denotes variables in the time-domain, $\tilde{\mathbf{w}}_\ell = [\tilde{w}_\ell[0], \cdots, \tilde{w}_\ell[K-1]]^T$ is the FIR input filter for the ℓ^{th} loudspeaker, $\tilde{\mathbf{h}}_{m\ell} = [\tilde{h}_{m\ell}[0], \cdots, \tilde{h}_{m\ell}[I-1]]^T$ is the room impulse response from the input of the ℓ^{th} loudspeaker to the output of the m^{th} microphone, and $\tilde{s}[n]$ is the n^{th} sample of the input audio signal. To determine the filters $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1^T, \cdots, \tilde{\mathbf{w}}_L^T]^T$, estimates of the room impulse responses are required. The typical approach in sound zone applications is to estimate the response from each loudspeaker independently [4], [9], [10]. Hereby, the problem can be divided into two stages: first, the room impulse response or the equivalent transfer functions are estimated at the microphone positions, and, secondly, these estimates are used to determine the filters $\tilde{\mathbf{w}}$.

To estimate a transfer function, the sound field generated by a loudspeaker is measured at a given microphone position. The measurement is usually repeated to suppress noise in the estimate. Assuming the observed pressure at microphone m due to loudspeaker ℓ to be a length N periodic sequence, the measured pressure samples for one period can be written as

$$\tilde{\mathbf{p}}_{m\ell}^{(r)} = \tilde{\mathbf{H}}_{m\ell} \tilde{\mathbf{s}} + \tilde{\mathbf{e}}_{m\ell}^{(r)}, \tag{2}$$

where (r) indicates the r^{th} repetition of the measurement, $\tilde{\mathbf{H}}_{m\ell}$ is the room impulse response represented as a convolutional matrix, $\tilde{\mathbf{e}}_{m\ell}^{(r)}$ is the noise present in the measurement, and $\tilde{\mathbf{s}}$ is a length N deterministic input sequence. To avoid aliasing in the convolutional matrix, it is assumed that N is longer than the impulse response of the room. With the assumption of a periodic input signal, $\tilde{\mathbf{H}}_{m\ell}$ is circulant and the pressure samples can be written as [29]

$$\tilde{\mathbf{p}}_{m\ell}^{(r)} = N^{-1} \mathbf{F}^H \operatorname{diag}(\mathbf{F}\tilde{\mathbf{h}}_{m\ell}) \mathbf{F}\tilde{\mathbf{s}} + \tilde{\mathbf{e}}_{m\ell}^{(r)}. \tag{3}$$

In the above, $\mathbf{F}$ is the N-by-N discrete Fourier transform (DFT) matrix as defined in [29] ($[\mathbf{F}]_{n,n'} = \exp^{(n-1)(n'-1)}(-2\pi j/N)$) and $(\cdot)^H$ denotes the Hermitian transpose. Recognizing $\operatorname{diag}(\mathbf{F}\tilde{\mathbf{h}}_{m\ell})\mathbf{F}\tilde{\mathbf{s}}$ as an entry wise multiplication between the vectors $\mathbf{F}\tilde{\mathbf{h}}_{m\ell}$ and $\mathbf{F}\tilde{\mathbf{s}}$, their ordering can be reversed to obtain

$$\tilde{\mathbf{p}}_{m\ell}^{(r)} = N^{-1} \mathbf{F}^H \operatorname{diag}(\mathbf{F}\tilde{\mathbf{s}}) \mathbf{F}\tilde{\mathbf{h}}_{m\ell} + \tilde{\mathbf{e}}_{m\ell}^{(r)}. \tag{4}$$

By multiplying both sides of the equation with the DFT matrix, the pressure samples are expressed in terms of the DFT coefficients

$$\mathbf{F}\tilde{\mathbf{p}}_{m\ell}^{(r)} = \operatorname{diag}(\mathbf{F}\tilde{\mathbf{s}}) \mathbf{F}\tilde{\mathbf{h}}_{m\ell} + \mathbf{F}\tilde{\mathbf{e}}_{m\ell}^{(r)}. \tag{5}$$

It is assumed that the samples in the noise sequence $\tilde{\mathbf{e}}_{m\ell}^{(r)}$, are drawn from a colored Gaussian distribution

$$\tilde{\mathbf{e}}_{m\ell}^{(r)} \sim \mathcal{N}(0, \tilde{\mathbf{C}}). \tag{6}$$

Multiplication with the DFT matrix is a linear transformation, hence, the transformed sequence is distributed as

$$\mathbf{F}\tilde{\mathbf{e}}_{m\ell}^{(r)} \sim \mathcal{CN}(0, \mathbf{\Lambda}), \tag{7}$$

where $\mathcal{CN}(\cdot, \cdot)$ denotes the complex Gaussian distribution. If we assume N is much larger than the effective auto-correlation time of the noise sequence, the covariance matrix $\mathbf{\Lambda}$ is asymptotically diagonal and related to $\tilde{\mathbf{C}}$ as $\mathbf{\Lambda} = \mathbf{F}^H \tilde{\mathbf{C}} \mathbf{F}/N$ [30]. From the linear properties of (5), the pressure observations are distributed as

$$\mathbf{F}\tilde{\mathbf{p}}_{m\ell}^{(r)} \sim \mathcal{CN}(\operatorname{diag}(\mathbf{F}\tilde{\mathbf{s}}) \mathbf{F}\tilde{\mathbf{h}}_{m\ell}, \mathbf{\Lambda}). \tag{8}$$

As the individual DFT coefficients are independent under these assumptions, a single element of (5) is expressed in the more compact form

$$p_{m\ell}^{(r)}(f_n) = s(f_n) h_{m\ell}(f_n) + e_{m\ell}^{(r)}(f_n), \tag{9}$$

where $f_n = n/N$ is the normalized frequency associated with each DFT coefficient. Now, assume the excitation signal consists of N_p pure-tones with frequencies on the DFT grid. If the DFT coefficients of interest are extracted and concatenated in a vector, we can write

$$\mathbf{p}_{m\ell}^{(r)} = \mathbf{S}\mathbf{h}_{m\ell} + \mathbf{e}_{m\ell}^{(r)}, \tag{10}$$

with

$$\mathbf{p}_{m\ell}^{(r)} = \begin{bmatrix} p_{m\ell}^{(r)}(f_1) & \cdots & p_{m\ell}^{(r)}(f_{N_p}) \end{bmatrix}^T \tag{11}$$

$$\mathbf{S} = \operatorname{diag}(\mathbf{s}), \quad \mathbf{s} = \begin{bmatrix} s(f_1) & \cdots & s(f_{N_p}) \end{bmatrix}^T \tag{12}$$

$$\mathbf{e}_{m\ell}^{(r)} = \begin{bmatrix} e_{m\ell}^{(r)}(f_1) & \cdots & e_{m\ell}^{(r)}(f_{N_p}) \end{bmatrix}^T. \tag{13}$$

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

The DFT coefficients of interest for a single measurement repetition are distributed as

$$\mathbf{p}_{m\ell}^{(r)} \sim \mathcal{CN}(\mathbf{S}\mathbf{h}_{m\ell}, \mathbf{\Sigma}_e), \tag{14}$$

where $\mathbf{\Sigma}_e = \text{diag}(\boldsymbol{\sigma}_e^2)$ are the corresponding N_p diagonal entries of $\mathbf{\Lambda}$.

Given R repeated measurements $\mathbf{p}_{m\ell} = [\mathbf{p}_{m\ell}^{(1)T}, \cdots, \mathbf{p}_{m\ell}^{(R)T}]^T$, the following approach was previously used in [28] to determine the transfer function as

$$\hat{h}_{m\ell}(f_n) = \frac{s^*(f_n)}{R|s(f_n)|^2} \sum_{r=1}^{R} p_{m\ell}^{(r)}(f_n), \tag{15}$$

where $(\cdot)^*$ denotes complex conjugation and $\hat{\cdot}$ highlights that $\hat{h}_{m\ell}(f_n)$ is an estimate of the noise-free $h_{m\ell}(f_n)$. However, in the presented work it is of interest to determine a probability distribution for each $\hat{h}_{m\ell}(f_n)$ to utilize the uncertainty inherently described in the distribution for regularizing the sound field control problem. In the appendix, it is shown how the probability distribution $g(\mathbf{h}_{m\ell})$ can be determined using variational Bayesian inference. This is used in sec. II-B, where the estimated probability distributions are included in the control problem to account for the uncertainty of the estimated transfer functions.

A. Reference Sound Field Control

For the reference sound field control, it is assumed that the estimated transfer functions are noise-free. Using estimates like (15), the reproduced sound pressure at the microphones is the summation of the filtered responses from the L loudspeakers and can be expressed as

$$\mathbf{p}_{\text{r}}(f_n) = \hat{\mathbf{H}}(f_n)\mathbf{w}(f_n) \tag{16}$$

$$\mathbf{p}_{\text{r}}(f_n) = \begin{bmatrix} p_{\text{r},1}(f_n) & \cdots & p_{\text{r},M}(f_n) \end{bmatrix}^T, \tag{17}$$

$$\hat{\mathbf{H}}(f_n) = \begin{bmatrix} \hat{h}_{11}(f_n) & \cdots & \hat{h}_{1L}(f_n) \\ \vdots & \ddots & \vdots \\ \hat{h}_{M1}(f_n) & \cdots & \hat{h}_{ML}(f_n) \end{bmatrix}, \tag{18}$$

$$\mathbf{w}(f_n) = \begin{bmatrix} w_1(f_n) & \cdots & w_L(f_n) \end{bmatrix}^T. \tag{19}$$

Sound zones can be created by optimizing cost functions consisting of different metrics related to this reproduced pressure. In the sound zones literature, two common metrics are the acoustic contrast and the reproduction error. The reproduction error for a given set of filters is

$$J_{\text{pm,r}}(\mathbf{w}(f_n)) = \|\mathbf{p}_t(f_n) - \hat{\mathbf{H}}(f_n)\mathbf{w}(f_n)\|_2^2, \tag{20}$$

where $\mathbf{p}_t(f_n) = [p_{t,1}(f_n), \cdots, p_{t,M}(f_n)]^T$ denotes a target sound field at the M microphone positions. The pressure matching solution is attained when minimizing $J_{\text{pm,r}}(\mathbf{w}(f_n))$ [15]. The acoustic contrast as an evaluation metric is defined as

$$\text{contrast}(f_n) = \frac{M_B^{-1} \sum_{m_B=1}^{M_B} |p_{\text{r},m_B}(f_n)|^2}{M_D^{-1} \sum_{m_D=1}^{M_D} |p_{\text{r},m_D}(f_n)|^2}, \tag{21}$$

where the subscripts $(\cdot)_B$ and $(\cdot)_D$ denote the bright and dark zone respectively. The acoustic contrast control solution is obtained by maximizing the cost-function

$$J_{\text{acc,r}}(\mathbf{w}(f_n)) = \frac{\|\hat{\mathbf{H}}_B(f_n)\mathbf{w}(f_n)\|_2^2}{\|\hat{\mathbf{H}}_D(f_n)\mathbf{w}(f_n)\|_2^2}, \tag{22}$$

as described in e.g. [13]. Minimizing (20) and maximizing (21) yields the reference PM and ACC filters [13], [15]

$$\mathbf{w}_{\text{pm,r}}(f_n) = \left(\hat{\mathbf{H}}(f_n)^H \hat{\mathbf{H}}(f_n)\right)^{-1} \hat{\mathbf{H}}(f_n)^H \mathbf{p}_t(f_n) \tag{23}$$

$$\mathbf{w}_{\text{acc,r}}(f_n) = \Theta\left((\hat{\mathbf{H}}_D(f_n)^H \hat{\mathbf{H}}_D(f_n))^{-1} \hat{\mathbf{H}}_B(f_n)^H \hat{\mathbf{H}}_B(f_n)\right) \tag{24}$$

with $\Theta(\cdot)$ denoting a function returning the eigenvector of its argument corresponding to the maximal eigenvalue.

B. Proposed Sound Field Control

If information about the marginal distribution of the transfer functions is available from (51), it can be incorporated in the cost functions. Let $\mathbf{G}(f_n)$ denote a matrix where all the entries are independent random variables, distributed according to (51). A common way to utilize the knowledge of the transfer function distributions is to augment the sound field control problems to optimize the expected objective [23]. Optimizing the expectation of the two sound field control problems, the cost-functions are rewritten as

$$J_{\text{pm,e}}(\mathbf{w}(f_n)) = \mathbb{E}\left\{ \|\mathbf{p}_t(f_n) - \mathbf{G}(f_n)\mathbf{w}(f_n)\|_2^2 \right\} \tag{25}$$

$$J_{\text{acc,e}}(\mathbf{w}(f_n)) = \mathbb{E}\left\{ \frac{\|\mathbf{G}_{\text{B}}(f_n)\mathbf{w}(f_n)\|_2^2}{\|\mathbf{G}_{\text{D}}(f_n)\mathbf{w}(f_n)\|_2^2} \right\}. \tag{26}$$

Minimizing the expectation in (25), a closed-form solution can be found as

$$\mathbf{w}_{\text{pm,e}}(f_n) = \left(\hat{\mathbf{G}}^H \hat{\mathbf{G}} + \sum_{m=1}^{M} \mathbf{\Sigma}_m \right)^{-1} \hat{\mathbf{G}}^H \mathbf{p}_t(f_n), \tag{27}$$

with $\hat{\mathbf{G}}$ being a matrix containing the expected value of each random variable in the entries of $\mathbf{G}(f_n)$, and σ_{ml}^2 being the corresponding variance from (53) used in

$$\mathbf{\Sigma}_m = \text{diag}\left(\begin{bmatrix} \sigma_{m1}^2 & \cdots & \sigma_{mL}^2 \end{bmatrix} \right). \tag{28}$$

As the numerator and denominator in (26) are independent in terms of the variance of the estimates, the expectation is taken separately. Thereby, it can be shown that (26) is maximized by

$$\mathbf{w}_{\text{acc,e}}(f_n) = \Theta\left(\left(\hat{\mathbf{G}}_D^H \hat{\mathbf{G}}_D + \sum_{m=1}^{M_D} \mathbf{\Sigma}_m \right)^{-1} \left(\hat{\mathbf{G}}_B^H \hat{\mathbf{G}}_B + \sum_{m=1}^{M_B} \mathbf{\Sigma}_m \right) \right). \tag{29}$$

The main difference between the proposed and reference solutions is the diagonal loadings introduced by the variances of the marginal distributions of the transfer function estimates. The diagonal loadings arise from the uncertainty in the measurements and thus automatically regularize the problems relative to the accuracy of the transfer function estimates.

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

III. Analysis of the pressure matching problem

To understand the challenge associated with creating sound zones at low frequencies in a room, we will here briefly look into the properties of the pressure matching problem. In this section, the sound field and transfer functions arise from solutions to partial differential equations and are expressed as continuous functions of space and the harmonic excitation frequency ω. Assuming noise-free transfer functions the PM problem is expressed from (20) as

$$\min_{F\{\tilde{\mathbf{w}}\}(\omega)} \|p_t(\bar{\mathbf{x}}_\mathrm{M},\omega) - \mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)F\{\tilde{\mathbf{w}}\}(\omega)\|_2^2. \tag{30}$$

where $\bar{\mathbf{x}}_\mathrm{M}$ denotes the M points where the sound field is observed, $\bar{\mathbf{x}}_\mathrm{L}$ denotes the L point source locations and $F\{\tilde{\mathbf{w}}\}(\omega)$ denotes the discrete-time Fourier transform of the filters, evaluated at ω. The properties of this problem depend on $\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)$, which is here related to the eigenfunctions of a room. At low frequencies, the sound field in a room due to harmonic excitation can be described in terms of an expansion in eigenfunctions (or room modes), each of which must satisfy the homogeneous Helmholtz equation [31]. The pressure at position $\mathbf{x}$ given a point source at $\mathbf{x}_0$ is

$$p_{\mathbf{x}_0}(\mathbf{x},\omega) = j\omega\rho G(\mathbf{x}|\mathbf{x}_0,\omega)q(\mathbf{x}_0,\omega), \tag{31}$$

where $j=\sqrt{-1}$, $q(\mathbf{x}_0)$ is the volume velocity emitted by the point source, and ρ is the density of air. Green's function, $G(\mathbf{x}|\mathbf{x}_0,\omega)$, relates a normalized volume velocity point source at $\mathbf{x}_0$ to the complex pressure at $\mathbf{x}$ and can be written as an eigenfunction expansion [31]

$$G(\mathbf{x}|\mathbf{x}_0,\omega) = \sum_{v=0}^{\infty} \mu_v(\omega)\psi_v(\mathbf{x})\psi_v(\mathbf{x}_0) \tag{32}$$

with

$$\mu_v(\omega) = \frac{1}{(\kappa_v^2 - (\omega/c)^2)}. \tag{33}$$

Here, c is the speed of sound and $\psi_v(\mathbf{x})$ is the v^{th} eigenfunction of the room[2] with eigenvalue κ_v^2 evaluated at point $\mathbf{x}$ in space. The sound pressure at $\mathbf{x}$ is the contribution from all point sources in the room,

$$p(\mathbf{x},\omega) = j\omega\rho \int_\Omega G(\mathbf{x}|\mathbf{x}_0,\omega)q(\mathbf{x}_0,\omega)d\mathbf{x}_0, \tag{34}$$

where the integral is taken over the domain of the room Ω.

For sound field control in a room, the sound field is generated by L point sources with individual volume velocities modified by the filters $\tilde{\mathbf{w}}$. Hence, (34) can be written as

$$p(\mathbf{x},\omega) = j\omega\rho \int_\Omega G(\mathbf{x}|\mathbf{x}_0,\omega) \sum_{\ell=1}^{L} F\{\tilde{\mathbf{w}}_\ell\}(\omega)\delta(\mathbf{x}_0-\mathbf{x}_\ell)d\mathbf{x}_0. \tag{35}$$

with $\delta(\cdot)$ being the Dirac delta and $\mathbf{x}_\ell$ the position of the ℓ^{th} point source. Assuming the microphone is not co-located with the point sources ($\mathbf{x}\neq\mathbf{x}_\ell$), that boundaries are not lossless, and that the amplitudes of the filters at ω are finite, the integrand is absolute integrable for each ℓ. Thereby, the order of integration and summation can be interchanged, and by substitution of (32) we obtain

$$p(\mathbf{x},\omega) = j\omega\rho \sum_{\ell=1}^{L} F\{\tilde{\mathbf{w}}_\ell\}(\omega) \sum_{v=0}^{\infty} \mu_v(\omega)\psi_v(\mathbf{x})\psi_v(\mathbf{x}_\ell). \tag{36}$$

From (36), it is seen that $\mu_v(\omega)$ reduce the influence of the eigenfunctions when the difference between ω^2 and $c^2\kappa_v^2$ is large. As such, the eigenfunctions are denoted as active when $\mu_v(\omega)$ is large (when the excitation frequency is close to the characteristic frequency of the eigenfunction).

In the pressure matching problem (30), the objective is to control the sound field in the room at M specific microphone positions. It is assumed that these positions represent the sound field within the zones and that it is not desired to explicitly control the sound field outside the zones. With the eigenfunction expansion of the sound field, the transfer function matrix in (30) can be written

$$\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega) = j\omega\rho\boldsymbol{\Psi}^H(\bar{\mathbf{x}}_\mathrm{M})\,\mathrm{diag}(\boldsymbol{\mu}(\omega))\boldsymbol{\Psi}(\bar{\mathbf{x}}_\mathrm{L}), \tag{37}$$

where

$$\boldsymbol{\Psi}^H(\bar{\mathbf{x}}_\mathrm{M}) = \begin{bmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_\infty(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \psi_1(\mathbf{x}_M) & \cdots & \psi_\infty(\mathbf{x}_M) \end{bmatrix} \tag{38}$$

$$\boldsymbol{\mu}(\omega) = \begin{bmatrix} \mu_1(\omega) & \cdots & \mu_\infty(\omega) \end{bmatrix}^T \tag{39}$$

$$\boldsymbol{\Psi}(\bar{\mathbf{x}}_\mathrm{L}) = \begin{bmatrix} \psi_1(\mathbf{x}_1) & \cdots & \psi_1(\mathbf{x}_L) \\ \vdots & \ddots & \vdots \\ \psi_\infty(\mathbf{x}_1) & \cdots & \psi_\infty(\mathbf{x}_L) \end{bmatrix}. \tag{40}$$

It is well known that the sensitivity of least-squares problems like (30) to perturbations in the target sound field and the transfer function matrix depends on the properties of $\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)$ [32]. The sensitivity is often evaluated in terms of the condition number of $\mathbf{H}^H(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)$, but if the eigenfunctions of the room are known it is possible to gain more insight into the problem. This is demonstrated with the following simulation example, where transfer functions from eight point sources are observed at 75 microphone positions in each of the two zones in a rectangular room, as sketched in Fig. 2. The eigenfunctions with characteristic frequency below 400 Hz are determined using the analytic expressions for Green's function as found in [26]. Further description of the simulated setup is given in section IV. For the simulated noise-free scenario, the condition number of $\mathbf{H}^H(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M},\bar{\mathbf{x}}_\mathrm{L},\omega)$ is plotted in Fig. 3. The condition number is high, especially at low frequencies, when compared against the suggested rule-of-thumb stating it should be in the interval 1000 to 5000 [33]. This generally indicates that the inverse problem (30) is sensitive to perturbations.

A more detailed understanding of the problem can be attained by analyzing how the microphone and source positions sample the active eigenfunctions. In the current example the number of microphones exceeds the number of sources, hence, the least-squares problem is overdetermined and thus no exact solution exists in general. However, closer examination reveals that the problem is almost underdetermined at low frequencies, which is the cause for the high condition number

[2]Note that the eigenfunctions here are scaled to be orthonormal, rather than the customary orthogonal with inner product equal to the volume of the room [26], [31].

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

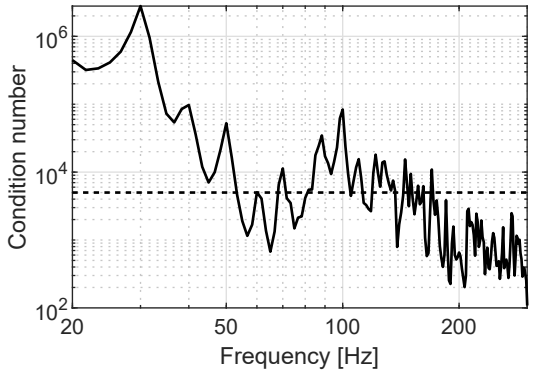


Fig. 3. Condition number of $\mathbf{H}^H(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)$ for the simulated setup sketched in fig. 2, having 150 microphones and 8 loudspeakers. (——): Condition number. (- - -): Reference limit of 5000 proposed in [33].

in Fig. 3. By observing that $\mathbf{\Psi}^H(\bar{\mathbf{x}}_M)$ and $\mathbf{\Psi}(\bar{\mathbf{x}}_L)$ are the eigenfunctions of the room evaluated at the microphone and source positions respectively, it is possible to analyze their contribution to $\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)$ independently. By truncating the number of eigenfunctions to V and separating $\mathrm{diag}(\boldsymbol{\mu}(\omega))$ in two diagonal matrices, (37) can be written as

$$\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega) = j\omega\rho\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}\left(\sqrt{\boldsymbol{\mu}_V(\omega)}\right) \mathrm{diag}\left(\sqrt{\boldsymbol{\mu}_V(\omega)}\right)\mathbf{\Psi}_V(\bar{\mathbf{x}}_L), \quad (41)$$

where the subscript V indicates truncation of (38)-(40) to order V.[3] With this separation it is of interest to investigate the degree to which the active eigenfunctions are different when observed by the microphones in the zones, at a given excitation frequency. Similarly, it is of interest to investigate whether each source can excite the active eigenfunctions differently from the other sources. The effective number of active eigenfunctions within the zones will be referred to as the observable degrees of freedom (DOF). In the same way, the effective number of active eigenfunctions which can be independently excited by the sources will be referred to as the controllable degrees of freedom. A mathematical definition of these DOF will follow after a brief motivation for this interpretation. In general, the eigenfunctions which are active at low frequencies change slowly as functions of the position in the room. As the frequency increases other eigenfunctions are active and these change more rapidly with the position in the room. The observable DOF might, therefore, be lower than the number of microphones in the room at low frequencies. As an example, consider two microphones sufficiently close together in the room. The pressure observed at each of the microphones would be nearly identical. Thus, only one DOF is observed not two.[4] Here, sufficiently close depends on both the physical distance and the active eigenfunctions. The interest in the observable and controllable DOF stems from the insight it provides to the properties of $\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)$, as will be discussed shortly.

[3]The truncation is justified by the negligible contribution from the eigenfunctions when ω^2 is far from $c^2\kappa_v^2$. The number of considered eigenfunctions is chosen to be larger than both the number of microphones and sources.

[4]A similar interpretation could be made for two sources in the room.

The chosen method for analyzing the observable and controllable DOF is through the singular value decomposition (SVD) of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ and $\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})\mathbf{\Psi}_V(\bar{\mathbf{x}}_L)$, independently. Both matrices can be approximated by a truncated SVD using the singular values and vectors corresponding to the first k singular values. By the Eckhart-Young theorem, the error of the rank-k truncated SVD approximation to a matrix $\mathbf{A}$ is $\|\mathbf{A} - \mathbf{A}_k\|_2 = s_{k+1}$, where s_{k+1} denotes the $(k+1)^{\text{th}}$ singular value of $\mathbf{A}$ [29]. The observable DOF is defined as the lowest rank of the truncated SVD of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ required to attain an approximation error below an application-specific threshold, relative to the maximal singular value of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$. The controllable DOF are similarly defined for the matrix $\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})\mathbf{\Psi}_V(\bar{\mathbf{x}}_L)$.

The normalized singular values of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ and $\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})\mathbf{\Psi}_V(\bar{\mathbf{x}}_L)$ are plotted in Fig. 4. If the application-specific threshold for observable and controllable DOF is set to a normalized approximation error of 0.1 (the dashed line in Fig. 4), it is seen that all the normalized singular values of $\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})\mathbf{\Psi}_V(\bar{\mathbf{x}}_L)$ are larger than 0.1. Thus, the controllable DOF are equal to the number of sources. However, several of the normalized singular values of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ are below 0.1, hence, the observable DOF are fewer than the number of microphones. Furthermore, it is seen that in the range 30-50 Hz the observable DOF are fewer than the controllable DOF. This is caused by few eigenfunctions being active in this frequency range, and the ones which are active change slowly throughout the room. Hence, as the microphones only evaluate the eigenfunctions in a small part of the room (see Fig. 2), it is reasonable that there are few observable DOF. The sources on the other hand are distributed throughout the entire room, and this spatial diversity increases the controllable DOF. As the frequency increases, a larger number of eigenfunctions are active (the modal density of the room increases [31]), and the active eigenfunctions change more rapidly across distance in the room. The effect is that the observable and controllable DOF increases, as observed from Fig. 4.

From this analysis, it is seen that $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\mathrm{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ is nearly rank-deficient at the lowest frequencies, which explains the ill-conditioning observed in Fig. 3. As this ill-conditioning is caused by having fewer observable DOF than controllable DOF, it is expected that regularization methods [32], would be effective to reduce the sensitivity caused by the ill-conditioning while retaining the ability to reproduce the target sound field in the zones, which can be described by the observable DOF. At higher frequencies, $\mathbf{H}^H(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)$ becomes well-conditioned but not all the observable DOF can be controlled. The expected outcome is that the sound field can be less accurately controlled at higher frequencies in this noise-free scenario. In the following sections, results are presented to show the consequence of $\mathbf{H}(\bar{\mathbf{x}}_\mathrm{M}, \bar{\mathbf{x}}_\mathrm{L}, \omega)$ being influenced by noise.

IV. Simulation Results

The influence of noise in the transfer function estimates is investigated using simulations where the noise-free transfer

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

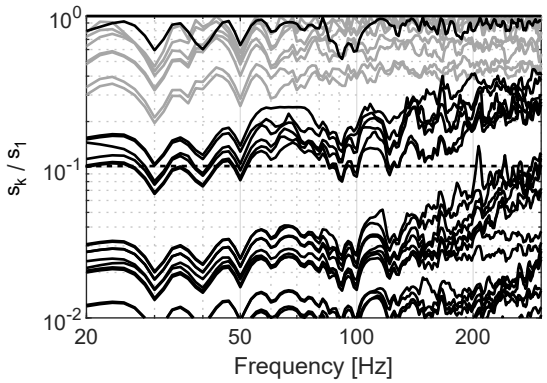


Fig. 4. Normalized singular values of $\mathbf{\Psi}_V^H(\bar{\mathbf{x}}_M)\text{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})$ (——) and $\text{diag}(\sqrt{\boldsymbol{\mu}_V(\omega)})\mathbf{\Psi}_V(\bar{\mathbf{x}}_L)$ (——). Example of threshold for determing the observable and controllable degrees of freedom (- - -).

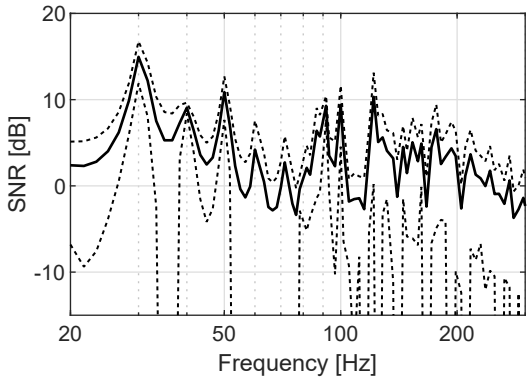


Fig. 5. Average signal level at the microphones, $\pm$ 1 standard deviation, relative to noise floor, when the source output is scaled for 0 dB free field SNR. Average and standard deviation are calculated across all the investigated 150×8 microphone loudspeaker pairs.

functions are known. Here a 5.5 m by 8.65 m by 2.7 m rectangular room with 0.6 s reverberation time (T_{60}) is simulated. The simulations are made using Green's function for a point source in a rectangular room [26], [31]. To emulate the measurement scenario, the room transfer functions are calculated from 8 point sources to 75 microphone positions (a 5×5 planar array in three different heights with 10 cm spacing between microphones[5]) in each zone, see the sketch in Fig. 2 for the layout. The corresponding room impulse responses (RIRs) are determined by the inverse DFT. The measurement procedure is emulated by convolving the RIRs with the excitation signal and adding uncorrelated white Gaussian noise. The excitation signal consists of multiple sinusoids with frequencies adjusted to be periodic with the measurement length N, so they are exactly on the analyzed DFT grid. The measurements are repeated eight times, and the marginal distributions for each transfer function bin is determined from (51) at convergence of the evidence lower bound (the relative change being below 10^{-9}).

The volume velocities of the point sources in the simulations are scaled to produce 70 dB SPL at a distance of 1 m in free field. A 2nd-order high-pass filter with cutoff frequency at 35 Hz is introduced to emulate the low frequency roll-off of a closed-box loudspeaker [34]. The noise added to the microphones is scaled to achieve a given SNR if the microphone had been positioned 1 m from the point source in free field and is denoted the free field SNR. The transfer functions of the room alter the frequency response, hence, the SNR observed at the microphones is fluctuating with the response of the room. The average signal to noise ratio $\pm$ 1 standard deviation across all the 150×8 microphone to point source pairs is plotted in Fig. 5 for a free field SNR of 0 dB. As an example, the transfer function estimates, (51) and (15), based on eight 0.6 s measurements are plotted in Fig. 6 along with the noise-free transfer function. It is seen that the two estimates are almost identical, hence, the main difference is the uncertainty inherently described by the Bayesian estimate.

[5]The microphone separation of 10 cm ensures the distance is less than 1/10 wavelength in the frequency range of interest.

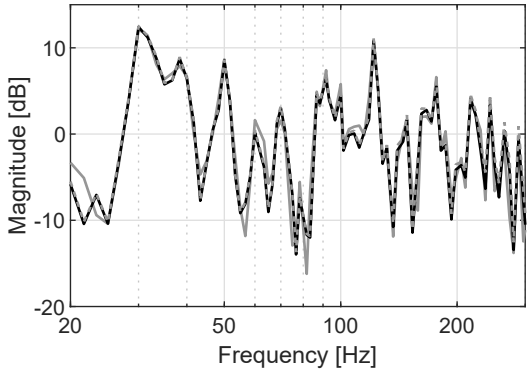


Fig. 6. Example of a transfer function estimated from 8 repeated measurements at 0 dB free field SNR. (——): True transfer function, (——): Bayesian estimate (51), (- - -) regular estimate (15).

A. Monte Carlo Results

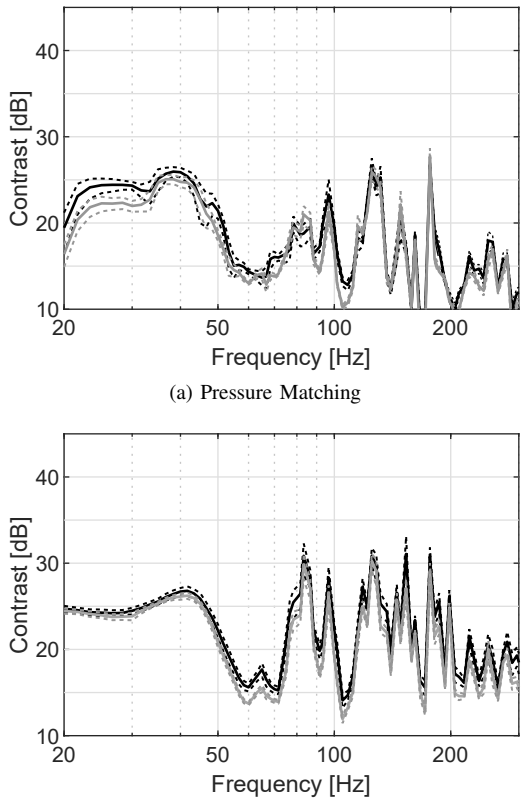
The effect of noise in the transfer function estimation is investigated through Monte Carlo simulations. The transfer functions are estimated from 8 measurements at a given free field SNR. These transfer functions are then used to calculate the sound field control filters as (23), (24), (27), and (29). The target sound field used in (27) and (23) is silence in the dark zone and the estimated transfer functions from the nearest loudspeaker in the bright zone.[6] The sound field control performance is evaluated as the contrast between the zones using the noise-free transfer functions. This procedure is repeated 100 times with new noise realizations, and the results are reported in terms of the average performance $\pm$ 1 standard deviation.

In Figs. 7 and 8, the contrast performance is shown for Monte Carlo simulations at free field SNR levels of 0 and 20 dB, respectively. Together with the average performance summarized in Table I, it is observed that the control based on the Bayesian estimates on average perform slightly worse than the reference methods but exhibits lower standard deviation.

[6]This target is chosen to ensure the target sound field in the bright zone is in the range of the estimated transfer function matrix ($\hat{\mathbf{H}}_\text{B}(f_n)$, $\hat{\mathbf{G}}_\text{B}$). Choosing the target as e.g. a plane impinging wave would imply adding room response equalization (see e.g. [35]) on top of the sound zones problem.

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

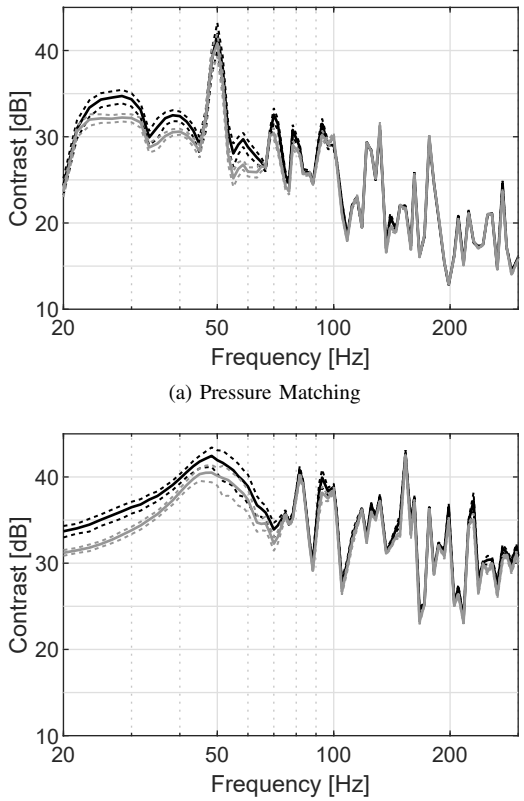
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing



(a) Pressure Matching

(b) Acoustic Contrast Control

Fig. 7. Average contrast (solid) $\pm$ 1 standard deviation (dashed) from 100 Monte Carlo simulations in a simulated rectangular room at 0 dB free field SNR. (——): $\mathbf{w}_{\text{pm,r}}$ (23), $\mathbf{w}_{\text{acc,r}}$ (24). (——): $\mathbf{w}_{\text{pm,e}}$ (27), $\mathbf{w}_{\text{acc,e}}$ (29)).



(a) Pressure Matching

(b) Acoustic Contrast Control

Fig. 8. Average contrast (solid) $\pm$ 1 standard deviation (dashed) from 100 Monte Carlo simulations in a simulated rectangular room at 20 dB free field SNR. (——): $\mathbf{w}_{\text{pm,r}}$ (23), $\mathbf{w}_{\text{acc,r}}$ (24). (——): $\mathbf{w}_{\text{pm,e}}$ (27), $\mathbf{w}_{\text{acc,e}}$ (29)).

This is expected as the white noise added to the simulated measurements is independent between microphones and sources. Thereby, the noise effectively regularizes the transfer function matrices in a similar manner as described in [22].

The contrast increases with the SNR, as expected from the estimated transfer function matrices becoming more accurate. It is also observed that at 0 dB free field SNR, the contrast performance from ACC is only slightly better than PM. However, at 20 dB free field SNR, the contrast performance using ACC is noticeably higher.

The efficiency of the methods can be compared based on the ℓ_2-norm of the filters required to create an average pressure of 94 dB SPL in the bright zone. This indicates how much energy is injected in the loudspeakers and how much sound is introduced outside the controlled zones. In Figs. 9, 10 and Table I, the norms are shown for free field SNRs of 0 and 20 dB. It is observed that the regularization based on the Bayesian estimates reduces the filter norms. The norms are also seen to increase with SNR as the regularization due to both inherent noise in the estimates and the uncertainty-based regularization decrease. Hence, the increased contrast at 20 dB free field SNR is achieved at the expense of increased energy consumption. Finally, it is seen that PM requires higher filter norms than ACC to produce the same average pressure in the bright zone, as expected from the literature [21].

TABLE I
SUMMARIZED RESULTS FROM FIG. 7-12 AVERAGED ACROSS FREQUENCY

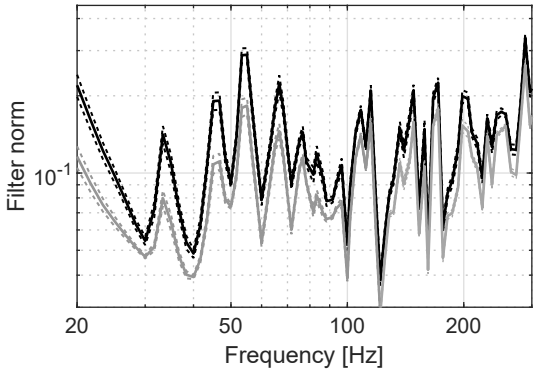
Method	0 dB free field SNR Contrast	Filter norm
$\mathbf{w}_{\text{pm,r}}$	17.4 ± 0.8 dB	$1.4 \times 10^{-1} \pm 7.3 \times 10^{-3}$
$\mathbf{w}_{\text{pm,e}}$	16.2 ± 0.6 dB	$1.1 \times 10^{-1} \pm 4.3 \times 10^{-3}$
$\mathbf{w}_{\text{acc,r}}$	21.2 ± 0.8 dB	$8.6 \times 10^{-2} \pm 3.5 \times 10^{-3}$
$\mathbf{w}_{\text{acc,e}}$	19.6 ± 0.5 dB	$7.7 \times 10^{-2} \pm 1.5 \times 10^{-3}$
	20 dB free field SNR Contrast	Filter norm
$\mathbf{w}_{\text{pm,r}}$	25.3 ± 0.5 dB	$2.3 \times 10^{-1} \pm 5.7 \times 10^{-3}$
$\mathbf{w}_{\text{pm,e}}$	24.5 ± 0.4 dB	$2.1 \times 10^{-1} \pm 3.6 \times 10^{-3}$
$\mathbf{w}_{\text{acc,r}}$	34.2 ± 0.5 dB	$1.4 \times 10^{-1} \pm 4.6 \times 10^{-3}$
$\mathbf{w}_{\text{acc,e}}$	33.1 ± 0.4 dB	$1.3 \times 10^{-1} \pm 2.2 \times 10^{-3}$
	Measured Results Contrast	Filter norm
$\mathbf{w}_{\text{pm,r}}$	21.2 dB	6.4×10^{-2}
$\mathbf{w}_{\text{pm,e}}$	21.6 dB	3.9×10^{-2}
$\mathbf{w}_{\text{acc,r}}$	27.7 dB	4.6×10^{-2}
$\mathbf{w}_{\text{acc,e}}$	28.4 dB	2.7×10^{-2}

V. EXPERIMENTAL RESULTS

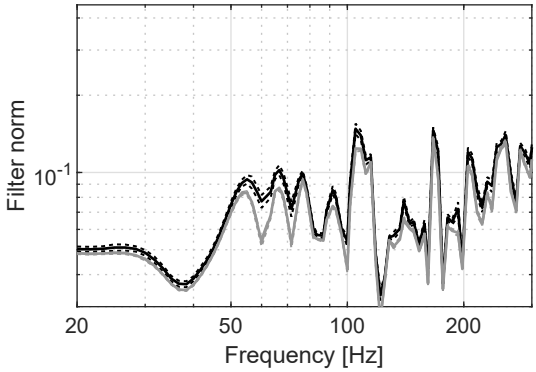
A series of in-situ measurements were conducted to validate the simulated results. The room was an office remodeled to resemble a living room. The shape of the room differs from the simulated rectangular room by having a raised ceiling, increasing the volume to 143 m^3. The layout of the 10" closed box loudspeakers and the 1/4" electret microphones were identical to the simulated results. The average measured (T_{20}) reverberation time in the frequency range of interest

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

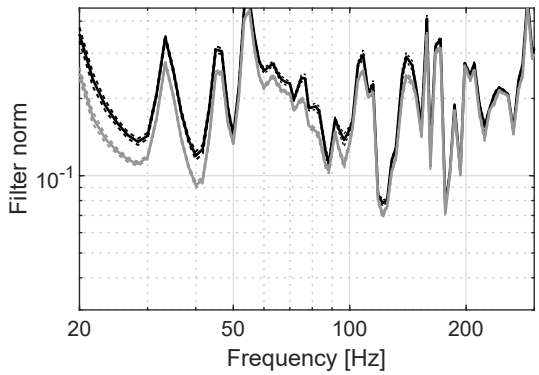


(a) Pressure Matching - Incoherent noise

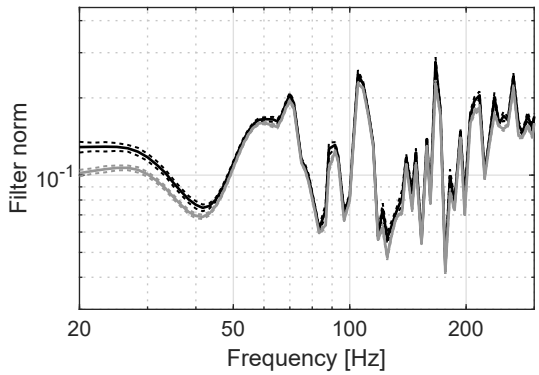


(b) Acoustic Contrast Control - Incoherent noise

Fig. 9. Average filter norms (solid) $\pm$ 1 standard deviation (dashed) from 100 Monte Carlo simulations in a simulated rectangular room at 0 dB free field SNR. (——): $\mathbf{w}_{\text{pm,r}}$ (23), $\mathbf{w}_{\text{acc,r}}$ (24). (——): $\mathbf{w}_{\text{pm,e}}$ (27), $\mathbf{w}_{\text{acc,e}}$ (29)).



(a) Pressure Matching



(b) Acoustic Contrast Control

Fig. 10. Average filter norm (solid) $\pm$ 1 standard deviation (dashed) from 100 Monte Carlo simulations in a simulated rectangular room at 20 dB free field SNR. (——): $\mathbf{w}_{\text{pm,r}}$ (23), $\mathbf{w}_{\text{acc,r}}$ (24). (——): $\mathbf{w}_{\text{pm,e}}$ (27), $\mathbf{w}_{\text{acc,e}}$(29)).

was 0.6 s. RME MADIface XT and M-32 AD / DA sound cards were used to measure the responses at 48 kHz sampling frequency before the results were downsampled to 1.2 kHz for further processing. The transfer functions were estimated from measurements at two excitation signal levels: one low for determining the sound zones filters, and one high for evaluating the results with little noise influence. Each measurement consisted of 10 repetitions of the 0.6 s periodic measurement sequence, where the first and last repetition were excluded from the processing. Before the measurements were commenced, the background noise power spectra for each of the 5×5 microphones in the planar array were estimated from 60 consecutive 0.6 s recordings of the noise. The average ratio of recorded signal power spectra to recorded background noise spectra is plotted in Fig. 11 for the measurements at low and high excitation signal levels. The evaluation measurements were conducted at 20 dB higher excitation level than the measurements used for calculating the filters, reaching maximum values of 82 dB SPL.

The contrast and filter norm results are plotted in Fig. 12 and summarized in Table I. As in the simulated results, the PM solutions provide reduced contrast at a higher filter norm. Contrary to the simulated results the contrast results for the expectation based cost-functions (27), (29) are slightly better than the reference solutions (23), (24). Furthermore, it is seen that the filter norms are significantly reduced by the explicit regularization.

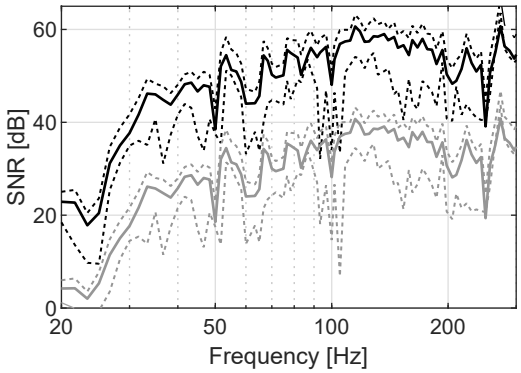


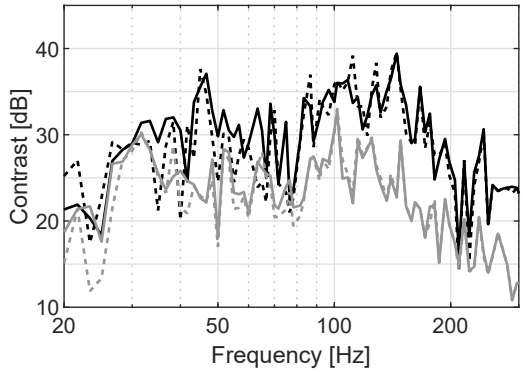
Fig. 11. Average ratio (solid) $\pm$ 1 standard deviation (dashed) of measured signal to measured background noise. (——): Measurements used to determine filters. (——): Measurements used to evaluate the results.

VI. Discussion

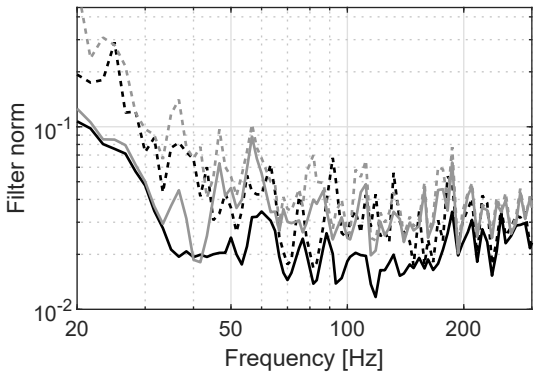
The slightly different relationship between proposed and reference methods in the experimental and simulated results suggests that the noise is different in the experimental results. One cause could be that the noise is partly correlated between microphones which is reasonable for acoustical noise at low frequencies. To support this argument, a Monte Carlo simulation was made where the added noise was perfectly correlated between the microphones. In Fig. 13 the contrast results are plotted for simulations with a free-field SNR of 20 dB. It is seen that the proposed and reference methods are

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

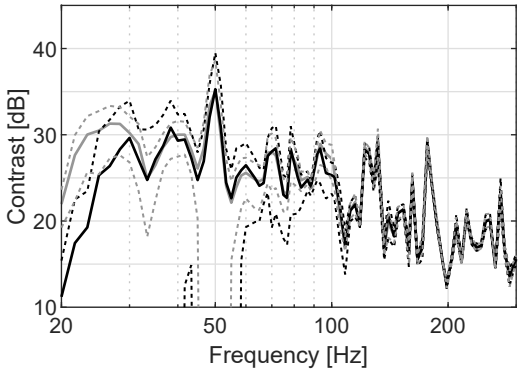


(a) Contrast

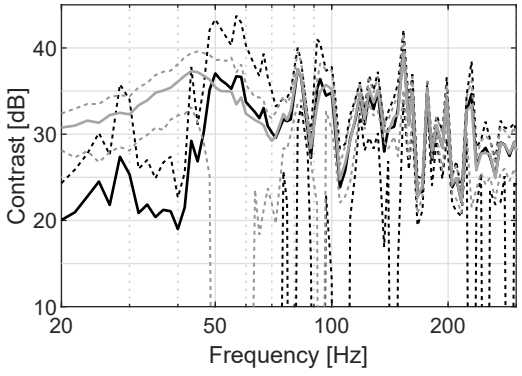


(b) Filter Norm

Fig. 12. Experimental results. Sound field control based on 8 repeated low SNR transfer function measurements, evaluated with a single high SNR measurement. (——): $\mathbf{w}_{\text{pm,e}}$ (27). (——): $\mathbf{w}_{\text{acc,e}}$ (29). (- - -): $\mathbf{w}_{\text{pm,r}}$ (23). (- - -): $\mathbf{w}_{\text{acc,r}}$ (24).



(a) Pressure Matching



(b) Acoustic Contrast Control

Fig. 13. Average contrast (solid) $\pm$ 1 standard deviation (dashed) from 100 Monte Carlo simulations with coherent noise in a simulated rectangular room at 20 dB free field SNR. (——): $\mathbf{w}_{\text{pm,r}}$ (23), $\mathbf{w}_{\text{acc,r}}$ (24). (——): $\mathbf{w}_{\text{pm,e}}$ (27), $\mathbf{w}_{\text{acc,e}}$ (29)).

different below 100 Hz, especially in terms of the standard deviation of the contrast. It is seen that the perfectly correlated noise does not inherently regularize the problem, hence, the particular noise sequence which is generated in each iteration of the Monte Carlo simulations has a great impact on the resulting contrast. Compared to the results in Fig. 7, 8, and 12, a much greater impact effect of the proposed methods is observed in Fig. 13. The perceptual difference between the reference and proposed methods are likely to be small for the uncorrelated noise simulations and the given experiment, due to the differences in contrast being mainly above 20 dB, in small parts of the investigated frequency range. However, for the correlated noise, the ± 1 standard deviation is clearly different between the reference and proposed methods. The added robustness increases the stability, and the lower range of the standard deviation are in some parts of the frequency range improved from below 10 dB to more than 20 dB. It is therefore argued, that the proposed methods have little negative effect when the noise is uncorrelated but can have a large possitive effect when the noise is correlated, even though the noise is modeled as uncorrelated in the Bayesian inference of the transfer functions.

Comparing the contrast performance across frequency in Figs. 7, 8, and 12, it is seen that the measured results follow a different trend than the simulated results. This change is due to a difference in the SNRs across frequency in the simulations

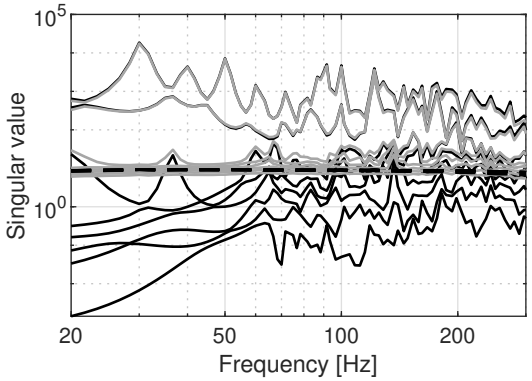


Fig. 14. Average singular values of $\hat{\mathbf{G}}^H\hat{\mathbf{G}}$ over 100 Monte Carlo trials at 0 dB free field SNR (——). (——): Singular values of the noise free transfer functions ($\mathbf{H}(f_n)^H\mathbf{H}(f_n)$). (- - -): Average singular values of the uncertainty based regularization $\sum_{m=1}^{M}\mathbf{\Sigma}_m$.

and the measurements. From Fig. 5 it is seen that the average SNR of the measurements has a slightly decreasing trend in the simulations. However, in Fig. 11 it is seen that the average SNR of the measured responses increases with frequency. Hereby, the accuracy of the estimated transfer functions over frequency is different for the simulated and measured results.

Besides the SNR, the frequency dependence of the contrast can be related to the eigenfunctions of the room. In Fig. 14 the singular values of the spatial correlation matrix of the Bayesian

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

estimates, $\hat{\mathbf{G}}^H\hat{\mathbf{G}}$, are plotted together with the noise-free spatial correlation matrix, $\mathbf{H}(f_n)^H\mathbf{H}(f_n)$, and the uncertainty based regularization for PM. It is seen that the added noise regularizes $\hat{\mathbf{G}}^H\hat{\mathbf{G}}$ and that the uncertainty corresponds to the added noise. At the lowest frequencies, only the first two singular values are correctly identified. However, as seen from Fig. 4 and 7, only few degrees of freedom (DOF) are observable in the zones at low frequencies, hence, it is possible to achieve more than 20 dB of contrast regardless. As the frequency increases, so does the observable DOF, as seen from both Fig. 4 and the diminishing ratio between largest and smallest singular value in Fig. 14. Thus, more loudspeakers are required to achieve a large contrast as frequency increases. However, as seen from Fig. 14 and 7 the additional loudspeakers are not useful if their added contribution cannot be resolved from the transfer functions due to the measurement noise. Thus, as the frequency increases it becomes increasingly important to accurately estimate the transfer functions in order to utilize all eight available loudspeakers for attaining a high contrast.

The results presented in this paper are limited to one room with a reverberation time of 0.6 s and a given layout of loudspeakers and microphones. However, the analysis presented can be interpreted for various reverberation times. Decreasing the reverberation time of the room decreases the maximal frequency dependent gain of each eigenfunction $\mu_v(\omega)$ and broadens the frequency span where it is active. This yields a greater frequency overlap between the eigenfunctions which increases both the observable and controllable DOF, as more eigenfunctions are relevant at any particular excitation frequency. At the same time, the decreased reverberation time also reduces the maximal amplification by the room. Thus, if the noise level is constant and the reverberation time is decreased, the maximal SNR (as plotted in Fig. 5) will decrease and the minimal SNR will increase. The result is that the contrast becomes more stable across frequency, as both the height of the peaks and the depth of the dips in the contrast performance decreases. On the other hand, if the reverberation time increases the opposite trend would be observed.

In this paper, the focus has been on introducing regularization based on the uncertainty in the estimated transfer functions. However, as can be inferred from Fig. 14, increasing the accuracy of the transfer functions also increases the condition number of the spatial correlation matrix. Therefore, having accurate measurements only improves the attainable contrast if the transfer functions do not change e.g. due to the temperature in the room changing. As the sound zones are assumed created with a feed-forward control system, the transfer functions are measured once and any later deviation from those estimates will degrade the contrast of the system. Thus, when the transfer function estimates become increasingly accurate, the sensitivity to other changes must be considered e.g. inspired by the procedure suggested in [23]. At the same time, even if the proposed method does not change the attained contrast significantly in the case of uncorrelated noise, it is valuable to know the uncertainty in the transfer functions as the inherent regularization in the noise reduces the sensitivity to other changes in the system.

VII. Conclusion

In this article, it was investigated how noise in in-situ transfer function estimates affects the achievable acoustic separation between sound zones in a room in the frequency range 20 - 300 Hz. By analyzing the way point sources radiate sound within a rectangular room the concepts of observable and controllable degrees of freedom were introduced. The conditioning of the transfer function matrix was related to the relationship between observable and controllable degrees of freedom. For the investigated setup, it was seen that when the controllable degrees of freedom exceed the observable degrees of freedom, it is possible to create more than 20 dB of separation even when the transfer functions are degraded by noise.

The automatic regularization based on the uncertainty of the Bayesian inference was shown to increase the robustness of the solution if the noise in the measurements is correlated between the microphones. If the noise at the microphones is uncorrelated, the influence of the regularization was shown to be small as the uncorrelated noise inherently regularizes the problem. Given high accuracy estimates of the transfer functions, further regularization might be required to increase the robustness towards e.g. time varying changes in the transfer functions.

Appendix A
Bayesian inference of the transfer function

With Bayesian inference it is possible to estimate the transfer functions and express the certainty of the estimate given the observed data. Throughout the appendix, the signal model from (10) - (14) is used, while the subscript $(\cdot)_{m\ell}$ and the frequency dependence (f_n) are ommitted for the ease of notation. The procedure for determining the marginal distribution for the transfer function is summarized in the end of the appendix.

It is assumed that the prior probability density function (pdf) of the transfer functions follows a complex Gaussian distribution with known mean and diagonal covariance matrix known up to a scalar, γ,

$$f(\mathbf{h}|\gamma) = \mathcal{CN}(\mathbf{h}; \boldsymbol{\mu}_h, \gamma\boldsymbol{\Sigma}_h). \tag{42}$$

The scalar is introduced to represent a frequency independent scaling of the covariance matrix $\boldsymbol{\Sigma}_h$. For tractability, the scalar is assumed to follow an inverse Gamma distribution

$$f(\gamma) = \mathcal{G}^{-1}(\gamma; a_\gamma, b_\gamma). \tag{43}$$

As it was assumed that the DFT coefficients of the noise sequence are independent, we can express the prior distribution for the noise variance as

$$f(\boldsymbol{\Sigma}_e) = \prod_{n=1}^{N_p} \mathcal{G}^{-1}(\sigma^2_{e,n}; a_n, b_n). \tag{44}$$

If the measurements are repeated R times, the data can be combined and used for estimating the transfer functions. The concatenation of the R observed data segments $\mathbf{p} = [\mathbf{p}^{(1)T}, \cdots, \mathbf{p}^{(R)T}]^T$ is distributed as

$$f(\mathbf{p}|\mathbf{h}, \boldsymbol{\Sigma}_e) = \mathcal{CN}(\mathbf{1}_R \otimes \mathbf{Sh}, \mathbf{I}_R \otimes \boldsymbol{\Sigma}_e), \tag{45}$$

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

where $\mathbf{1}_R = [1, \cdots, 1]^T \in \mathbb{R}^R$, $\mathbf{I}_R \in \mathbb{R}^{R\times R}$ is the identity matrix, and $\otimes$ denotes the Kronecker product. From Bayes theorem, the posterior distribution of the latent variables is given as

$$f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e|\mathbf{p}) = \frac{f(\mathbf{p}|\mathbf{h}, \mathbf{\Sigma}_e) f(\mathbf{h}|\gamma) f(\gamma) f(\mathbf{\Sigma}_e)}{f(\mathbf{p})}. \tag{46}$$

Because the prior of the transfer functions is independent of the measurement noise, it is not possible to factor the joint distribution $f(\mathbf{p}, \mathbf{h}, \gamma, \mathbf{\Sigma}_e)$ into independent components of noise variance $\sigma^2_{e,n}$ and variance of the transfer function prior $\sigma^2_{h,n}$. Therefore, we cannot analytically determine the marginal distribution of interest, $f(\mathbf{h}|\mathbf{p})$. This can be overcome by enforcing a dependence through the g-prior [36], [37] or by directly estimating the properties of the marginal distribution through Markov chain Monte Carlo simulations [38]. In the presented work, however, the marginal distribution is approximated analytically through variational inference. In this framework, a family of pdfs are used to approximate the joint posterior distribution. Here, it is assumed that the joint probability can be factorized as

$$f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e|\mathbf{p}) \approx g(\mathbf{h}, \gamma, \mathbf{\Sigma}_e) = g_1(\mathbf{h}) g_2(\mathbf{\Sigma}_e) g_3(\gamma). \tag{47}$$

The approximation is optimized in terms of the Kullback-Leibler divergence between $g(\mathbf{h}, \gamma, \mathbf{\Sigma}_e)$ and $f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e, \mathbf{p})$. In [38], it is shown that the optimal choices for $g_1(\mathbf{h})$, $g_2(\mathbf{\Sigma}_e)$, and $g_3(\gamma)$ are

$$\ln(g_1(\mathbf{h})) = \mathop{\mathbb{E}}_{i\neq 1}\{\ln(f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e, \mathbf{p}))\} + \text{const.} \tag{48}$$

$$\ln(g_2(\mathbf{\Sigma}_e)) = \mathop{\mathbb{E}}_{i\neq 2}\{\ln(f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e, \mathbf{p}))\} + \text{const.} \tag{49}$$

$$\ln(g_3(\gamma)) = \mathop{\mathbb{E}}_{i\neq 3}\{\ln(f(\mathbf{h}, \gamma, \mathbf{\Sigma}_e, \mathbf{p}))\} + \text{const.}, \tag{50}$$

where $\mathop{\mathbb{E}}_{i\neq 1}\{\cdot\}$ denotes the expectation with respect to the distributions different from $g_1(\mathbf{h})$, i.e. $g_2(\mathbf{\Sigma}_e)$ and $g_3(\gamma)$. As the expression for $g_1(\mathbf{h})$ depends on the expectation with respect to $g_2(\mathbf{\Sigma}_e)$ and $g_3(\gamma)$ and vice versa, a consensus estimate must be determined iteratively as shown in the procedure overview at the end of the appendix. It can be shown that

$$g_1(\mathbf{h}) = \mathcal{CN}\left(\mathbf{h}; \boldsymbol{\mu}_{\mathbf{h}|\mathbf{p}}, \mathbf{\Sigma}_{\mathbf{h}|\mathbf{p}}\right) \tag{51}$$

$$\mathbf{\Sigma}_{\mathbf{h}|\mathbf{p}} = \text{diag}\left(\begin{bmatrix}\sigma^2_{h|\mathbf{p},1} & \cdots & \sigma^2_{h|\mathbf{p},N_p}\end{bmatrix}\right) \tag{52}$$

$$\sigma^2_{h|\mathbf{p},n} = \left(|s_n|^2 R \mathop{\mathbb{E}}_{g_2}\{\sigma^{-2}_{e,n}\} + \mathop{\mathbb{E}}_{g_3}\{\gamma^{-1}\}\sigma^{-2}_{h,n}\right)^{-1} \tag{53}$$

$$\boldsymbol{\mu}_{\mathbf{h}|\mathbf{p}} = \mathbf{\Sigma}_{\mathbf{h}|\mathbf{p}}\left(R\mathbf{S}^H \mathop{\mathbb{E}}_{g_2}\{\mathbf{\Sigma}_e^{-1}\}\mathbf{S}\langle\mathbf{p}\rangle + \mathop{\mathbb{E}}_{g_3}\{\gamma^{-1}\}\mathbf{\Sigma}_h^{-1}\boldsymbol{\mu}_h\right), \tag{54}$$

$$g_2(\mathbf{\Sigma}_e) = \prod_{n=1}^{N_p} \mathcal{G}^{-1}(\sigma^2_{e,n}; \alpha_n, \beta_n) \tag{55}$$

$$\alpha_n = a_n + R \tag{56}$$

$$\beta_n = b_n + R\left(\langle|p_n|^2\rangle + \mathop{\mathbb{E}}_{g_1}\{|h_n|^2\}|s_n|^2 - \mathop{\mathbb{E}}_{g_1}\{h_n^*\}s_n^*\langle p_n\rangle - \langle p_n^*\rangle s_n \mathop{\mathbb{E}}_{g_1}\{h_n\}\right), \tag{57}$$

and

$$g_3(\gamma) = \mathcal{G}^{-1}(\gamma; \alpha_\gamma, \beta_\gamma) \tag{58}$$

$$\alpha_\gamma = a_\gamma + N \tag{59}$$

$$\beta_\gamma = b_\gamma + \sum_{n=1}^{N_p} \sigma^{-2}_{h,n}\left(|\mu_{h,n}|^2 + \mathop{\mathbb{E}}_{g_1}\{|h_n|^2\} - \mathop{\mathbb{E}}_{g_1}\{h_n^*\}\mu_{h,n} - \mu^*_{h,n}\mathop{\mathbb{E}}_{g_1}\{h_n\}\right). \tag{60}$$

The notation $\langle\cdot\rangle$ is used to indicate the average across the R measurements.

From the properties of the inverse gamma distribution we have

$$\mathop{\mathbb{E}}_{g_2}\{\sigma^{-2}_{e,n}\} = \frac{\alpha_n}{\beta_n} \tag{61}$$

$$\mathop{\mathbb{E}}_{g_3}\{\gamma^{-1}\} = \frac{\alpha_\gamma}{\beta_\gamma} \tag{62}$$

and it can be shown that

$$\mathop{\mathbb{E}}_{g_1}\{|h_n|^2\} = |\mu_{h|\mathbf{p},n}|^2 + \sigma^2_{h|\mathbf{p},n}. \tag{63}$$

Hereby, it is possible to determine the distributions $g_1(\mathbf{h})$, $g_2(\mathbf{\Sigma}_e)$, and $g_3(\gamma)$ by iteratively updating the estimates. The convergence is determined from the evidence lower bound [38]

$$\begin{aligned}\text{ELBO} = &\mathop{\mathbb{E}}_{g_1,g_2}\{\ln(f(\mathbf{p}|\mathbf{h}, \mathbf{\Sigma}_e)\} + \mathop{\mathbb{E}}_{g_1,g_3}\{\ln(f(\mathbf{h}|\gamma))\} \\ &+ \mathop{\mathbb{E}}_{g_3}\{\ln(f(\gamma))\} + \mathop{\mathbb{E}}_{g_2}\{\ln(f(\mathbf{\Sigma}_e))\} + \mathbb{H}(g_1) + \mathbb{H}(g_2) + \mathbb{H}(g_3),\end{aligned} \tag{64}$$

with $\mathbb{H}(g_1)$, $\mathbb{H}(g_2)$, and $\mathbb{H}(g_3)$ denoting the entropies of $g_1(\mathbf{h})$, $g_2(\mathbf{\Sigma}_e)$, and $g_3(\gamma)$, respectively. The change in evidence lower bound can be evaluated between each cycle of updates and the iterations are stopped when the change is below a predefined threshold. The convergence is guaranteed because the ELBO is convex with respect to each of the factors $g_1(\mathbf{h})$, $g_2(\mathbf{\Sigma}_e)$, and $g_3(\gamma)$ [38]. The components of (64) are given as

$$\begin{aligned}\mathop{\mathbb{E}}_{g_1,g_2}\{\ln(f(\mathbf{p}|\mathbf{h}, \mathbf{\Sigma}_e))\} = &-RN_p\ln(\pi) - R\text{tr}(\mathop{\mathbb{E}}_{g_2}\{\ln(\mathbf{\Sigma}_e)\}) \\ &- R\sum_{n=1}^{N_p}\mathop{\mathbb{E}}_{g_2}\{\sigma^{-2}_{e,n}\}\left(\langle|p_n|^2\rangle + \mathop{\mathbb{E}}_{g_1}\{|h_n|^2\}|s_n|^2 - \mathop{\mathbb{E}}_{g_1}\{h_n^*\}s_n^*\langle p_n\rangle - \langle p_n^*\rangle s_n\mathop{\mathbb{E}}_{g_1}\{h_n\}\right)\end{aligned} \tag{65}$$

$$\begin{aligned}\mathop{\mathbb{E}}_{g_1}\{\ln(f(\mathbf{h}|\gamma))\} = &-N_p\ln(\pi) - \text{tr}(\ln(\mathbf{\Sigma}_h)) - N_p\mathop{\mathbb{E}}_{g_3}\{\ln(\gamma)\} \\ &-\mathop{\mathbb{E}}_{g_3}\{\gamma^{-1}\}\sum_{n=1}^{N_p}\mathop{\mathbb{E}}_{g_1}\{|h_n|^2\} + |\mu^2_{h,n} - \mathop{\mathbb{E}}_{g_1}\{h_n^*\}\mu_{h,n} - \mu^*_{h,n}\mathop{\mathbb{E}}_{g_1}\{h_n\}\end{aligned} \tag{66}$$

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

TABLE II
ROOM PARAMETER RANGES FOR MONTE CARLO SIMULATIONS FOR DETERMINING A PRIOR DISTRIBUTION FOR THE TRANSFER FUNCTIONS IN A ROOM.

Parameter	Min. value	Max. value
Length	3 m	10 m
Width	3 m	10 m
Height	2.5 m	3.5 m
Temperature	20°	30°
T_{60}	0.1 s	0.7 s

$$\mathop{\mathbb{E}}_{g_2}\{\ln(f(\mathbf{\Sigma}_e))\} = \sum_{n=1}^{N_p} a_n \ln(b_n) - \ln(\Gamma(a_n)) - (a_n+1) \mathop{\mathbb{E}}_{g_2}\{\ln(\sigma_{e,n}^2)\} - b_n \mathop{\mathbb{E}}_{g_2}\{\sigma_{e,n}^{-2}\} \tag{67}$$

$$\mathop{\mathbb{E}}_{g_2}\{\ln(f(\gamma))\} = a_\gamma \ln(b_\gamma) - \ln(\Gamma(a_\gamma)) - (a_\gamma+1) \mathop{\mathbb{E}}_{g_3}\{\ln(\gamma)\} - b_\gamma \mathop{\mathbb{E}}_{g_3}\{\gamma^{-1}\} \tag{68}$$

$$\mathbb{H}(g_1) = \sum_{n=1}^{N_p} \ln(\pi e \sigma^2_{h|\mathbf{p},n}) \tag{69}$$

$$\mathbb{H}(g_2) = \sum_{n=1}^{N_p} \alpha_n + \ln(\beta_n \Gamma(\alpha_n)) - (\alpha_n+1)F(\alpha_n) \tag{70}$$

$$\mathbb{H}(g_3) = \alpha_\gamma + \ln(\beta_\gamma \Gamma(\alpha_\gamma)) - (\alpha_\gamma+1)F(\alpha_\gamma). \tag{71}$$

In the above, $\Gamma(\cdot)$ and $F(\cdot)$ represent the gamma and digamma functions, respectively.

A. Selection of Prior Hyperparameters

In the previous section, the hyper parameters $\boldsymbol{\mu}_h$, $\mathbf{\Sigma}_h$, a_γ, b_γ, $a_{e,n}$, and $b_{e,n}$ were unspecified. The scalar γ and the noise level are unknown prior to the measurements, hence a_γ, b_γ, $a_{e,n}$ and $b_{e,n}$ should be chosen to reflect this with a vague prior (e.g. $a_\gamma = b_\gamma = a_{e,n} = b_{e,n} = 10^{-6}$), such that the observed data is weighted much higher than the prior knowledge.

The prior distribution for the transfer function is assumed to be a complex Gaussian. To specify the hyperparameters of the distribution, it is assumed that the measurement will be made in a rectangular room, where the sound field adheres to the structure introduced in (31) for Green's function in rectangular rooms [31]. In this case, a Monte Carlo simulation is performed for rectangular rooms of various dimensions, temperatures, source-microphone positions, and reverberation times. For each simulation, a value is drawn uniformly from each of the intervals listed in Table II. The prior distribution is then determined from the sample mean and variance of a Monte Carlo simulation of 100,000 iterations.

B. Procedure for Bayesian inference of transfer function

With the equations and the hyper parameters defined, the procedure for estimating the marginal distribution for the transfer function $g_1(\mathbf{h})$ is summarized as the following steps:

1: Measure pressure response R times to acquire $\mathbf{p}$
2: Initialize hyper parameters: $a_\gamma, b_\gamma, a_{e,n}, b_{e,n}$
3: **while** $\Delta_{\text{ELBO}} >$ tolerance **do**
4: Update: $g_1(\mathbf{h})$ given $g_2(\mathbf{\Sigma}_e)$ and $g_3(\gamma)$
5: Update: $g_2(\mathbf{\Sigma}_e)$ given $g_1(\mathbf{h})$ and $g_3(\gamma)$
6: Update: $g_3(\gamma)$ given $g_1(\mathbf{h})$ and $g_2(\mathbf{\Sigma}_e)$
7: Update: ELBO given $g_1(\mathbf{h})$, $g_2(\mathbf{\Sigma}_e)$, and $g_3(\gamma)$
8: $\Delta_{\text{ELBO}} \leftarrow \| \frac{\text{ELBO} - \text{ELBO}_{\text{prev}}}{\text{ELBO}} \|$
9: $\text{ELBO}_{\text{prev}} \leftarrow \text{ELBO}$
10: **end while**
11: Return: $\boldsymbol{\mu}_{\mathbf{h}|\mathbf{p}}, \mathbf{\Sigma}_{\mathbf{h}|\mathbf{p}}$ describing $g_1(\mathbf{h})$.

REFERENCES

[1] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio. Eng. Soc.*, vol. 45, no. 9, pp. 685–701, September 1997.
[2] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 81–91, March 2015.
[3] M. Jones and S. J. Elliott, "Personal audio with multiple dark zones," *J. Acoust. Soc. Am.*, vol. 6, no. 124, pp. 3497–3506, December 2008.
[4] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, June 2013.
[5] J.-W. Choi, "Subband optimization for acoustic contrast control," in *Proc. of the 22nd International Congress on Sound and Vibration (ICSV22), Florence, Italy, 12-16 July*, 2015.
[6] M. Christoph and M. Kronlachner, "Improvement of personal sound zones by individual delay compensation," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jul 2016.
[7] M. Olsen and M. B. Møller, "Sound zones: on the effect of ambient temperature variations in feed-forward systems," in *Audio Engineering Society Convention 142*, May 2017.
[8] M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, "The effect of reverberation on personal audio devices," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2654–2663, 2014.
[9] F. Olivieri, M. Shin, F. M. Fazi, P. A. Nelson, and P. Otto, "Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.
[10] M. Olik, J. Francombe, P. Coleman, P. J. B. Jackson, M. Olsen, M. Møller, R. Mason, and S. Bech, "A comparative performance study of sound zoning methods in a reflective environment," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.
[11] M. B. Møller and M. Olsen, "Sound zones: on performance prediction of contrast control methods," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, July 2016.
[12] F. M. Heuchel, D. C. Nozal, and F. T. Agerkvist, "Sound field control for reduction of noise from outdoor concerts," in *Audio Engineering Society Convention 145*, Oct 2018.
[13] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, 2002.
[14] J. Rämö, S. Bech, and S. H. Jensen, "Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 153–163, 2018.
[15] M. Poletti, "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*, Oct 2008.
[16] F. Olivieri, F. M. Fazi, P. A. Nelson, M. Shin, S. Fontana, and L. Yue, "Theoretical and experimental comparative analysis of beamforming methods for loudspeaker arrays under given performance constraints," *J. Sound and Vibration*, vol. 373, pp. 302–324, July 2016.
[17] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1869–1878, November 2015.
[18] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2992–3000, November 1993.
[19] X. Ma, P. J. Hegarty, J. A. Pedersen, and J. J. Larsen, "Impact of loudspeaker nonlinear distortion on personal sound zones," *J. Acoust. Soc. Am.*, vol. 143, no. 1, pp. 51–59, 2018.

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2019.2921151, IEEE/ACM Transactions on Audio, Speech, and Language Processing

[20] J.-Y. Park, M.-H. Song, J.-H. Chang, and Y.-H. Kim, "Performance degradation due to transfer function errors in acoustic brightness and contrast control: sensitivity analysis," in *Proc. of the 20th International Congress on Acoustics (ICA 2010)*, 2010.
[21] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, 2014.
[22] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language processing*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.
[23] Q. Zhu, P. Coleman, M. Wu, and J. Yang, "Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling," *J. Audio Eng. Soc.*, vol. 65, no. 6, pp. 460–473, 2017.
[24] ——, "Robust personal audio reproduction based on acoustic transfer function modelling," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Jun 2016.
[25] ——, "Robust reproduction of sound zones with local sound orientation," *J. Acoust. Soc. Am.*, vol. 142, no. 1, pp. EL118–EL122, 2017.
[26] F. Jacobsen and P. M. Juul, *Fundamentals of General Linear Acoustics*. Wiley, 2013.
[27] *Orientering fra Miljøstyrelsen - Lavfrekvent støj, infralyd og vibrationer i eksternt miljø*. Miljø- og Energiministeriet. Miljøstyrelsen, 1997.
[28] M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, "On the influence of transfer function noise on low frequency pressure matching for sound zones," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018, pp. 331–335.
[29] G. H. Golub and C. F. V. Loan, *Matrix Computations, 4th edt.* Johns Hopking University Press, 2013.
[30] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, 1993.
[31] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968.
[32] P. C. Hansen, *Discrete inverse problems: insights and algorithms*. Society for Industrial and Applied Mathematics, 2010.
[33] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *J. Acoust. Soc. Am.*, vol. 100, no. 3, pp. 1584–1593, September 1996.
[34] W. M. Leach, *Introduction to Electroacoustics and Audio Amplifier Design, 3rd edt.* Kendall/Hunt, 2003.
[35] S. Cecchi, A. Carini, and S. Spors, "Room response equalization - a review," *Applied Sciences*, vol. 8, no. 1, 2018.
[36] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for bayesian variable selection," *J. Amer. Statist. Assoc.*, vol. 103, pp. 410–423, March 2008.
[37] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, and S. H. Jensen, "Bayesian model comparison with the g-prior," *IEEE Transactions on Signal Processing*, vol. 62, no. 1, pp. 225–238, Jan 2014.
[38] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.



Martin Bo Møller (S'17) received the M.Sc. degree in engineering acoustics from the Technical University of Denmark in 2011 and was employed by Bang & Olufsen A/S (B&O). Since 2013, he has been studying part-time towards the Ph.D. degree, sponsored by B&O, at the Department of Electronic Systems at Aalborg University, Denmark. His research interests include sound field control, loudspeaker array processing, as well as loudspeaker and room interaction.



Jesper Kjær Nielsen (S'12–M'13) received the M.Sc (Cum Laude) and Ph.D. degrees in electrical engineering with a specialisation in signal processing from Aalborg University, Denmark, in 2009 and 2012, respectively. From 2012 to 2016, he was with the Department of Electronic Systems, Aalborg University, as an industrial postdoctoral researcher (12-15) and as a non-tenured associate professor (15-16). Bang & Olufsen A/S (B&O) was the industrial partner in these four years. Jesper is currently with the Audio Analysis Lab, Aalborg University, in a three year position as an assistant professor in Statistical Signal Processing. He is part-time employed by B&O.

Jesper has been a Visiting Scholar in the Signal Processing and Communications Laboratory, University of Cambridge in 2009 and at the Department of Computer Science, University of Illinois at Urbana-Champaign in 2011. Moreover, he has been a guest researcher in the Signal & Information Processing Lab at TU Delft in 2014. His research interests include spectral estimation, (sinusoidal) parameter estimation, microphone array processing, as well as statistical and Bayesian methods for signal processing.



Efren Fernandez-Grande Efren Fernandez-Grande is an Associate Professor in the Acoustic Technology group, Department of Electrical Engineering at the Technical University of Denmark. He received the BSc degree in telecommunication engineering from the Polytechnic University of Madrid (UPM) in 2006, and MSc and PhD degrees in acoustics from the Technical University of Denmark (DTU) in 2008 and 2012 respectively. He obtained a postdoctoral fellowship from the Danish Council for Independent Research (2013-2015), was Assistant Professor at DTU between 2014 and 2017, and is Associate Professor since 2017. He is a VILLUM Young Investigator fellow (project "Large-scale acoustic holography", 2018-2023). His research interests are in the areas of acoustic signal processing, sound field analysis and sensor array processing, with a particular focus in room acoustics, sound radiation, acoustic holography and spatial audio. He teaches several courses in physical acoustics and signal processing, and since 2018 is the head of the MSc in Engineering Acoustics at DTU. He has been a visiting researcher at the Korea Advanced Institute of Science and Technology (KAIST), the Institut Langevin, The University of California San Diego (UCSD) and the LAUM Universite du Maine. He is a member of the Technical Committee in Audio Signal Processing of the EAA, and serves as Associate Editor for the Journal of the Acoustical Society of America, in the Acoustic Signal Processing section.



Søren Krarup Olesen Søren Krarup Olesen received the M.Sc. degree in 1991 in electrical engineering with specialisation in acoustics, and Ph.D. degree in 2001 with focus on binaural techniques and signal processing – both at Aalborg University, Denmark. He currently works as an associate professor at Aalborg University with teaching obligations, in particular programming, acoustical transducers, digital signal processing and electronics. Apart from programming, Søren has gained experience in the design of computer controlled listening experiments. Research interests include acoustical room simulation, 3D-sound and direction estimation of arrival of sound.

2329-9290 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.